PATENT COOPERATION TREATY

From the INTERNATIONAL BUREAU

PCT

NOTIFICATION OF ELECTION

(PCT Rule 61.2)

To:

Assistant Commissioner for Patents
United States Patent and Trademark Office
Box PCT
Washington, D.C.20231
ETATS-UNIS D'AMERIQUE

in its capacity as elected Office

Date of mailing (day/month/year)	
22 March 2000 (22.03.00)	
International application No. PCT/US99/15931	**Applicant's or agent's file reference** BB1174
International filing date (day/month/year) 14 July 1999 (14.07.99)	**Priority date** (day/month/year) 17 July 1998 (17.07.98)
Applicant CAHOON, Rebecca, E. et al	

1. The designated Office is hereby notified of its election made:

[X] in the demand filed with the International Preliminary Examining Authority on:

07 February 2000 (07.02.00)

[ ] in a notice effecting later election filed with the International Bureau on:

2. The election [X] was

[ ] was not

made before the expiration of 19 months from the priority date or, where Rule 32 applies, within the time limit under Rule 32.2(b).

The International Bureau of WIPO 34, chemin des Colombettes 1211 Geneva 20, Switzerland	Authorized officer Beate Giffo-Schmitt
Facsimile No.: (41-22) 740.14.35	Telephone No.: (41-22) 338.83.38

PATENT COOPERATION TREATY

PCT

INTERNATIONAL SEARCH REPORT

(PCT Article 18 and Rules 43 and 44)

Applicant's or agent's file reference	**FOR FURTHER ACTION** see Notification of Transmittal of International Search Report (Form PCT/ISA/220) as well as, where applicable, item 5 below.	
BB1174		
International application No.	International filing date *(day/month/year)*	(Earliest) Priority Date *(day/month/year)*
PCT/US 99/ 15931	14/07/1999	17/07/1998
Applicant		
E.I. DU PONT DE NEMOURS AND COMPANY et al.		

This International Search Report has been prepared by this International Searching Authority and is transmitted to the applicant according to Article 18. A copy is being transmitted to the International Bureau.

This International Search Report consists of a total of 3 sheets.

[X] It is also accompanied by a copy of each prior art document cited in this report.

1. **Basis of the report**

   a. With regard to the **language**, the international search was carried out on the basis of the international application in the language in which it was filed, unless otherwise indicated under this item.

   [ ] the international search was carried out on the basis of a translation of the international application furnished to this Authority (Rule 23.1(b)).

   b. With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, the international search was carried out on the basis of the sequence listing :

   [X] contained in the international application in written form.

   [ ] filed together with the international application in computer readable form.

   [ ] furnished subsequently to this Authority in written form.

   [ ] furnished subsequently to this Authority in computer readble form.

   [ ] the statement that the subsequently furnished written sequence listing does not go beyond the disclosure in the international application as filed has been furnished.

   [ ] the statement that the information recorded in computer readable form is identical to the written sequence listing has been furnished

2. [ ] **Certain claims were found unsearchable** (See Box I).

3. [ ] **Unity of invention is lacking** (see Box II).

4. With regard to the **title,**

   [X] the text is approved as submitted by the applicant.

   [ ] the text has been established by this Authority to read as follows:

5. With regard to the **abstract,**

   [X] the text is approved as submitted by the applicant.

   [ ] the text has been established, according to Rule 38.2(b), by this Authority as it appears in Box III. The applicant may, within one month from the date of mailing of this international search report, submit comments to this Authority.

6. The figure of the **drawings** to be published with the abstract is Figure No. 1

   [X] as suggested by the applicant.

   [ ] because the applicant failed to suggest a figure.

   [ ] because this figure better characterizes the invention.

   [ ] None of the figures.

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 99/15931

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N15/54 C12N15/82 C12N9/12 C12N5/10 C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12N C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	REITH M ET AL: "Two amino-acid biosynthetic genes are encoded on the plastid genome of the red alga Porphyra umbilicalis." CURRENT GENETICS, (1993 JAN) 23 (1) 59-65. , XP002118025 cited in the application the whole document ---	1-6
A	MCKAY G AND SHARGOOL P.: "Purification and characterization of N-acetylglutamate 5-phosphotransferase from pea (Pisum sativum) cotyledons" BIOCHEMICAL JOURNAL, vol. 195, no. 1, 1981, pages 71-81, XP002118026 see the whole document; for claim 7 see esp. p.73 r. col. --- -/--	1-7

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
7 October 1999	21/10/1999

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Kania, T

1

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
A	EP 0 261 627 A (KYOWA HAKKO KOGYO KK) 30 March 1988 (1988-03-30) the whole document -----	1-6

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/US 99/15931

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 0261627 A	30-03-1988	JP 1994602 C	22-11-1995
		JP 7028749 B	05-04-1995
		JP 63079597 A	09-04-1988
		DE 3785530 A	27-05-1993
		DE 3785530 T	18-11-1993
		US 5017482 A	21-05-1991

PATENT COOPERATION TREATY

RECEIVED

SEP 29 2000

PATENT RECORDS CENTER

From th
INTERNATIONAL PRELIMINARY EXAMINING AUTHORITY

To:

E.I. DU PONT DE NEMOURS AND COMPANY
Legal/Patent Records Center
1007 Market Street
Wilmington, Delaware 19898
ETATS-UNIS D'AMERIQUE

KL

PCT

NOTIFICATION OF TRANSMITTAL OF THE INTERNATIONAL PRELIMINARY EXAMINATION REPORT
(PCT Rule 71.1)

Date of mailing *(day/month/year)*	20.09.2000
Applicant's or agent's file reference BB1174	**IMPORTANT NOTIFICATION**

International application No.	International filing date *(day/month/year)*	Priority date *(day/month/year)*
PCT/US99/15931	14/07/1999	17/07/1998

Applicant
E.I. DU PONT DE NEMOURS AND COMPANY et al.

1. The applicant is hereby notified that this International Preliminary Examining Authority transmits herewith the international preliminary examination report and its annexes, if any, established on the international application.

2. A copy of the report and its annexes, if any, is being transmitted to the International Bureau for communication to all the elected Offices.

3. Where required by any of the elected Offices, the International Bureau will prepare an English translation of the report (but not of any annexes) and will transmit such translation to those Offices.

4. **REMINDER**

The applicant must enter the national phase before each elected Office by performing certain acts (filing translations and paying national fees) within 30 months from the priority date (or later in some Offices) (Article 39(1)) (see also the reminder sent by the International Bureau with Form PCT/IB/301).

Where a translation of the international application must be furnished to an elected Office, that translation must contain a translation of any annexes to the international preliminary examination report. It is the applicant's responsibility to prepare and furnish such translation directly to each elected Office concerned.

For further details on the applicable time limits and requirements of the elected Offices, see Volume II of the PCT Applicant's Guide.

TRB NOTED

Name and mailing address of the IPEA/	Authorized officer
European Patent Office D-80298 Munich Tel. +49 89 2399 - 0 Tx: 523656 epmu d Fax: +49 89 2399 - 4465	Vullo, C Tel.+49 89 2399-8061

Form PCT/IPEA/416 (July 1992)

17 JAN 2001

PATENT COOPERATION TREATY

PCT

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

(PCT Article 36 and Rule 70)

Applicant's or agent's file reference	FOR FURTHER ACTION	See Notification of Transmittal of International Preliminary Examination Report (Form PCT/IPEA/416)
BB1174		
International application No.	International filing date *(day/month/year)*	Priority date *(day/month/year)*
PCT/US99/15931	14/07/1999	17/07/1998
International Patent Classification (IPC) or national classification and IPC C12N15/54		
Applicant E.I. DU PONT DE NEMOURS AND COMPANY et al.		

1. This international preliminary examination report has been prepared by this International Preliminary Examining Authority and is transmitted to the applicant according to Article 36.

2. This REPORT consists of a total of 5 sheets, including this cover sheet.

   ☐ This report is also accompanied by ANNEXES, i.e. sheets of the description, claims and/or drawings which hav been amended and are the basis for this report and/or sheets containing rectifications made before this Authority (see Rule 70.16 and Section 607 of the Administrative Instructions under the PCT).

   These annexes consist of a total of sheets.

3. This report contains indications relating to the following items:

   - I ☒ Basis of the report
   - II ☐ Priority
   - III ☐ Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
   - IV ☐ Lack of unity of invention
   - V ☒ Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations suporting such statement
   - VI ☒ Certain documents cited
   - VII ☐ Certain defects in the international application
   - VIII ☒ Certain observations on the international application

Date of submission of the demand	Date of completion of this report
07/02/2000	20.09.2000
Name and mailing address of the international preliminary examining authority: European Patent Office D-80298 Munich Tel. +49 89 2399 - 0 Tx: 523656 epmu d Fax: +49 89 2399 - 4465	Authorized officer Heimann-Pohl, B Telephone No. +49 89 2399 8713

Form PCT/IPEA/409 (cover sheet) (January 1994)

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

International application No. PCT/US99/15931

I. Basis of the r port

1. This report has been drawn on the basis of (*substitute sheets which have been furnished to the receiving Office in response to an invitation under Article 14 are referred to in this report as "originally filed" and are not annexed to the report since they do not contain amendments.*):

Description, pages:

1-27 as originally filed

Claims, No.:

1-7 as originally filed

Drawings, sheets:

1 as originally filed

2. The amendments have resulted in the cancellation of:

☐ the description, pages:
☐ the claims, Nos.:
☐ the drawings, sheets:

3. ☐ This report has been established as if (some of) the amendments had not been made, since they have been considered to go beyond the disclosure as filed (Rule 70.2(c)):

4. Additional observations, if necessary:

V. Reasoned statement under Article 35(2) with regard t novelty, inventiv step or industrial applicability; citations and explanations supporting such statement

1. Statement

Novelty (N)	Yes:	Claims	1-7
	No:	Claims	
Inventive step (IS)	Yes:	Claims	1-6
	No:	Claims	7
Industrial applicability (IA)	Yes:	Claims	1-7
	No:	Claims	

2. Citations and explanations

see separate sheet

VI. Certain documents cited

1. Certain published documents (Rule 70.10)

and / or

2. Non-written disclosures (Rule 70.9)

see separate sheet

VIII. Certain observations on the international application

The following observations on the clarity of the claims, description, and drawings or on the question whether th claims are fully supported by the description, are made:

see separate sheet

1). The present application relates to the cloning and the sequences of cDNAs encoding N-acetylglutamate kinase from higher plants, maize, rice, wheat and soybean.

2). Prior Art

D1: REITH M ET AL: 'Two amino-acid biosynthetic genes are encoded on the plastid genome of the red alga Porphyra umbilicalis:' CURRENT GENETICS, (1993 JAN) 23 (1) 59-65. , XP002118025 cited in the application
D2: MCKAY G AND SHARGOOL P.: 'Purification and characterization of N-acetylglutamate 5-phosphotransferase from pea (Pisum sativum) cotyledons' BIOCHEMICAL JOURNAL, vol. 195, no. 1, 1981, pages 71-81, XP002118026
D3: EP-A-0 261 627 (KYOWA HAKKO KOGYO KK) 30 March 1988 (1988-03-30)

D1 discloses the cloning of ALS and N-acetylglutamate kinase coding sequences from red alga P. umbilicales plastids. The N-acetylglutamate kinase coding sequence shows an identity of 56.7 % on the protein level to the present maize sequence.

D2 discloses the purification and characterization of N-acetylglutamate kinase from pea and further relates to inhibitor studies with purified enzyme.

D3 discloses recombinant corynebacterium and brevibacterium strains transformed with N-acetylglutamate kinase coding sequence from E. coli for the production of arginine.

3). Novelty (Box V)

None of the above cited documents discloses the specific sequences and the chimeric genes or transformed host cells comprising the chimeric genes of claims 1-6. Therefore the subject matter of claims 1-6 is novel.

Furthermore, a method to identify compounds which inhibit recombinant N-

acetylglutamate kinase (claim 7) is also not disclosed in these documents.

4). Inventive Step (Box V)

Although it was known from D2 that N-acetylglutamate kinase exist in pea and possibly other higher plants, it cannot be taken from said document that specifically maize, rice, wheat and soybean would encode such an enzyme. Therefore the subject matter of claims 1-6 seems to involve an inventive step (Art. 33 (3) PCT).

The method of claim 7 is not restricted to N-acetylglutamate kinase from maize, rice, wheat and soybean, since it is not linked to the subject matter of claims 1-6. Said method differs from the method disclosed in D2 only by the use of a recombinant enzyme. However, using any prior art N- acetylglutamate kinase, e.g. D1 or D3 which are not excluded from claim 7, in the method of D2 cannot be called inventive. Thus the method of claim 7 lacks an inventive step (Art. 33 (3) PCT).

5). Clarity (Box VIII)

In claim 1 the term "encoding a functional enzyme" is unclear because this term does not define whether this function should be the one of N-acetylglutamate kinase or a different one. An illustrative example for a switch of a function can be taken from the abstract on page 18 of "Oncogenes and Growth Control 2000" (EMBL, SALK, EMBO Conference, 13.- 17. May 2000 EMBL Heidelberg) with regard to the presence or absence of the amino acids KTS (document annexed to this International Preliminary Examination Report).
Thus claim 1 lacks clarity, contrary to the requirements of Art. 6 PCT and the PCT International Preliminary Examination Guidelines, Chapter III-4.2.
Likewise, in claim 2 the term "functionally" lacks clarity.

PATENT COOPERATION TREATY

PCT

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

(PCT Article 36 and Rule 70)

Applicant's or agent's file reference	FOR FURTHER ACTION	See Notification of Transmittal of International Preliminary Examination Report (Form PCT/IPEA/416)
BB1174		
International application No.	International filing date *(day/month/year)*	Priority date *(day/month/year)*
PCT/US99/15931	14/07/1999	17/07/1998
International Patent Classification (IPC) or national classification and IPC C12N15/54		
Applicant E.I. DU PONT DE NEMOURS AND COMPANY et al.		

1. This international preliminary examination report has been prepared by this International Preliminary Examining Authority and is transmitted to the applicant according to Article 36.

2. This REPORT consists of a total of 5 sheets, including this cover sheet.

   ☐ This report is also accompanied by ANNEXES, i.e. sheets of the description, claims and/or drawings which have been amended and are the basis for this report and/or sheets containing rectifications made before this Authority (see Rule 70.16 and Section 607 of the Administrative Instructions under the PCT).

   These annexes consist of a total of sheets.

3. This report contains indications relating to the following items:

   - I ☒ Basis of the report
   - II ☐ Priority
   - III ☐ Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
   - IV ☐ Lack of unity of invention
   - V ☒ Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations suporting such statement
   - VI ☒ Certain documents cited
   - VII ☐ Certain defects in the international application
   - VIII ☒ Certain observations on the international application

Date of submission of the demand	Date of completion of this report
07/02/2000	20.09.2000
Name and mailing address of the international preliminary examining authority: European Patent Office D-80298 Munich Tel. +49 89 2399 - 0 Tx: 523656 epmu d Fax: +49 89 2399 - 4465	Authorized officer Heimann-Pohl, B Telephone No. +49 89 2399 8713

Form PCT/IPEA/409 (cover sheet) (January 1994)

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

International application No. PCT/US99/15931

I. Basis of the report

1. This report has been drawn on the basis of (*substitute sheets which have been furnished to the receiving Office in response to an invitation under Article 14 are referred to in this report as "originally filed" and are not annexed to the report since they do not contain amendments.*):

Description, pages:

1-27 as originally filed

Claims, No.:

1-7 as originally filed

Drawings, sheets:

1 as originally filed

2. The amendments have resulted in the cancellation of:

☐ the description, pages:
☐ the claims, Nos.:
☐ the drawings, sheets:

3. ☐ This report has been established as if (some of) the amendments had not been made, since they have been considered to go beyond the disclosure as filed (Rule 70.2(c)):

4. Additional observations, if necessary:

1). The present application relates to the cloning and the sequences of cDNAs encoding N-acetylglutamate kinase from higher plants, maize, rice, wheat and soybean.

2). Prior Art

D1: REITH M ET AL: 'Two amino-acid biosynthetic genes are encoded on the plastid genome of the red alga Porphyra umbilicalis.' CURRENT GENETICS, (1993 JAN) 23 (1) 59-65. , XP002118025 cited in the application
D2: MCKAY G AND SHARGOOL P.: 'Purification and characterization of N-acetylglutamate 5-phosphotransferase from pea (Pisum sativum) cotyledons' BIOCHEMICAL JOURNAL, vol. 195, no. 1, 1981, pages 71-81, XP002118026
D3: EP-A-0 261 627 (KYOWA HAKKO KOGYO KK) 30 March 1988 (1988-03-30)

D1 discloses the cloning of ALS and N-acetylglutamate kinase coding sequences from red alga P. umbilicales plastids. The N-acetylglutamate kinase coding sequence shows an identity of 56.7 % on the protein level to the present maize sequence.

D2 discloses the purification and characterization of N-acetylglutamate kinase from pea and further relates to inhibitor studies with purified enzyme.

D3 discloses recombinant corynebacterium and brevibacterium strains transformed with N-acetylglutamate kinase coding sequence from E. coli for the production of arginine.

3). Novelty (Box V)

None of the above cited documents discloses the specific sequences and the chimeric genes or transformed host cells comprising the chimeric genes of claims 1-6. Therefore the subject matter of claims 1-6 is novel.

Furthermore, a method to identify compounds which inhibit recombinant N-

acetylglutamate kinase (claim 7) is also not disclosed in these documents.

4). Inventive Step (Box V)

Although it was known from D2 that N-acetylglutamate kinase exist in pea and possibly other higher plants, it cannot be taken from said document that specifically maize, rice, wheat and soybean would encode such an enzyme. Therefore the subject matter of claims 1-6 seems to involve an inventive step (Art. 33 (3) PCT).

The method of claim 7 is not restricted to N-acetylglutamate kinase from maize, rice, wheat and soybean, since it is not linked to the subject matter of claims 1-6. Said method differs from the method disclosed in D2 only by the use of a recombinant enzyme. However, using any prior art N- acetylglutamate kinase, e.g. D1 or D3 which are not excluded from claim 7, in the method of D2 cannot be called inventive. Thus the method of claim 7 lacks an inventive step (Art. 33 (3) PCT).

5). Clarity (Box VIII)

In claim 1 the term "encoding a functional enzyme" is unclear because this term does not define whether this function should be the one of N-acetylglutamate kinase or a different one. An illustrative example for a switch of a function can be taken from the abstract on page 18 of "Oncogenes and Growth Control 2000" (EMBL, SALK, EMBO Conference, 13.- 17. May 2000 EMBL Heidelberg) with regard to the presence or absence of the amino acids KTS (document annexed to this International Preliminary Examination Report).
Thus claim 1 lacks clarity, contrary to the requirements of Art. 6 PCT and the PCT International Preliminary Examination Guidelines, Chapter III-4.2.
Likewise, in claim 2 the term "functionally" lacks clarity.

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 :
C12N 15/54, 15/82, 9/12, 5/10, C12Q 1/68 | A1

(11) International Publication Number: WO 00/04168

(43) International Publication Date: 27 January 2000 (27.01.00)

(21) International Application Number: PCT/US99/15931

(22) International Filing Date: 14 July 1999 (14.07.99)

(30) Priority Data:
60/093,209 17 July 1998 (17.07.98) US

(71) Applicant (*for all designated States except US*): E.I. DU PONT DE NEMOURS AND COMPANY [US/US]; 1007 Market Street, Wilmington, DE 19898 (US).

(72) Inventors; and
(75) Inventors/Applicants (*for US only*): CAHOON, Rebecca, E. [US/US]; 2331 West 18th Street, Wilmington, DE 19806 (US). GUTTERIDGE, Steven [US/US]; 4 Austin Road, Wilmington, DE 19810 (US). LEE, Jian–Ming [CN/US]; Apartment 203, 3120 Naamans Road, Wilmington, DE 19810 (US).

(74) Agent: MAJARIAN, William, R.; E.I. du Pont de Nemours and Company, Legal Patent Record Center, 1007 Market Street, Wilmington, DE 19898 (US).

(81) Designated States: AE, AL, AU, BA, BB, BG, BR, CA, CN, CU, CZ, EE, GD, GE, HR, HU, ID, IL, IN, IS, JP, KP, KR, LC, LK, LR, LT, LV, MG, MK, MN, MX, NO, NZ, PL, RO, SG, SI, SK, SL, TR, TT, UA, US, UZ, VN, YU, ZA, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.

(54) Title: ORNITHINE BIOSYNTHESIS ENZYMES

```
                    *                                                    +
SEQ ID NO:11  MSSTQDYIGE----------------------------------------------------
SEQ ID NO:02  MLLTKPYLSNSLLPVPSPPPSGPTLSSNHASPLAAPTCR-RSRLRISATSTAAPSPSSAA
SEQ ID NO:04  MLLAKPHLSSSSF-LPSTRVSSPAPGPNHAKPIAASPAP-RRCLRLAVTSAAAPAASSAE
SEQ ID NO:06  MMAG----AAKTLTNLCPSFPFPTKPQNQLTTSHAFPSTRLRHRAISAVANAAQPPLAAA
SEQ ID NO:08  MLLTKPH---PALTLPSASLPNPNLKAARVRPLASSAPHGRRGLRV---SASSSSLAPAQ
SEQ ID NO:12  MXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXAX
              1                                                          60

                +     **++*** **+** * *++ **********+++*+++** *+*+++++*+ **
SEQ ID NO:11  ---EAATRVKILSEALPYIQHFAGRTVVVKYGGAAMKDSNLKDKVIRDIVFMASVGIRPV
SEQ ID NO:02  AATASLSRVDVLSEALPFIQRFKGKTVVVKYGGAAMKSPELQASVIRDLVLLSCVGLRPV
SEQ ID NO:04  AA-AALSRVDVLSEALPFIQRFKGKTVVVKYGGAAMKSPELQASVIRDLVLLSCVGLRPV
SEQ ID NO:06  TATEGQYRVDVLSESLPFIQKFRGKTIVVKYGGAAMKSPELQASVINDLVLLSCVGLRPV
SEQ ID NO:08  AASAALNRVDVLSEALPFIQRFKGKTVVVKYGGAAMKSPELQASVIRDLVLLSCVGLRPV
SEQ ID NO:12   AXXXXXRVDVLSEXLPFIQXFXGKTXVVKYGGAAMKSPELQASVIXDLVLLSCVGLXPV
              61                                                         120

              +**********+** +      *+ *****                    * +**   *  *
SEQ ID NO:11  VVHGGGPEINTWLDKVGIEPQFKDGLRVTDAATMDIVEMVLVGRVNKSLVNLINQAGGKA
SEQ ID NO:02  LVHGGGPEINSWLLRVGVEPQFRDGLRVTDALTMEVVEMVLVGKVNKNLVSLINIAGGTA
SEQ ID NO:04  LVHGGGPEINSWLLRVGVEPQFRNGLRVTDALNMEVVEMVLVRKVNKELLSLIKLPGGSA
SEQ ID NO:06  LVHGGGPEINSWLGRLNIPAVFRDGLRVTDADTMEIVSMVLVGKVNKTLVSLINKAGATA
SEQ ID NO:08  LVHGGGPEINSWLQRVGVXPQFRNGLRVTXXXXXXXXXXXXXXXXXXKQLLSLIRPAGTTA
SEQ ID NO:12  LVHGGGPEINSWLXRXXXXXXFRXGLRVTDAXXMEXVXMVLVXKVNKXLXSLIXXXGXXA
              121                                                        180
```

(57) Abstract

This invention relates to an isolated nucleic acid fragment encoding an N–acetylglutamate kinase. The invention also relates to the construction of a chimeric gene encoding all or a portion of the N–acetylglutamate kinase, in sense or antisense orientation, wherein expression of the chimeric gene results in production of altered levels of the N–acetylglutamate kinase in a transformed cell host.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

TITLE

ORNITHINE BIOSYNTHESIS ENZYMES

This application claims the benefit of U.S. Provisional Application No. 60/093,209, filed July 17, 1998.

FIELD OF THE INVENTION

This invention is in the field of plant molecular biology. More specifically, this invention pertains to nucleic acid fragments encoding N-acetyl glutamate kinase in plants and seeds.

BACKGROUND OF THE INVENTION

Ornithine is converted into arginine in the urea cycle. Intermediaries in the ornithine biosynthesis pathway are important in other steps of this cycle. Amino acid N-acetyl transferase (EC 2.3.1.1) catalyzes the first reaction in a pathway that leads to the synthesis of ornithine from L-glutamate giving N-acetylglutamate as its intermediary product. Carbamoyl phosphate synthase I, the mitochondrial enzyme that catalyzes the first committed step of the urea cycle, is allosterically activated by N-acetyl glutamate. The rate of urea production by the liver is, in fact, correlated with the N-acetylglutamate concentration. Increased urea synthesis is required when amino acid breakdown rates increase, generating excess nitrogen that must be extracted. Increase in these breakdown rates are signaled by an increase in glutamate concentration through transamination reaction. This situation, in turn, causes an increase in N-acetylglutamate synthesis, stimulating carbamoyl phosphate synthetase and the entire urea cycle.

N-acetyl glutamate kinase (EC 2.7.2.8) catalyzes the conversion of N-acetyl-L-glutamate and ATP into N-acetyl-L-glutamate-5-phosphate and ADP. N-acetylglutamate kinase is encoded by the argB locus in bacteria and *Synechocystis*. This enzyme has been characterized at the molecular level in the red alga *Porphyra umbilicalis*, a member of the protista kingdom (Reith, M., Munholland, J. (1993) *Curr. Genet. 23*:59-65), but it has yet to be described in higher plants.

SUMMARY OF THE INVENTION

The instant invention relates to isolated nucleic acid fragments encoding N-acetyl glutamate kinase. Specifically, this invention concerns an isolated nucleic acid fragment encoding an N-acetyl glutamate kinase and an isolated nucleic acid fragment that is substantially similar to an isolated nucleic acid fragment encoding an N-acetyl glutamate kinase. In addition, this invention relates to a nucleic acid fragment that is complementary to the nucleic acid fragment encoding N-acetyl glutamate kinase.

An additional embodiment of the instant invention pertains to a polypeptide encoding all or a substantial portion of an N-acetyl glutamate kinase.

In another embodiment, the instant invention relates to a chimeric gene encoding an N-acetyl glutamate kinase, or to a chimeric gene that comprises a nucleic acid fragment that

is complementary to a nucleic acid fragment encoding an N-acetyl glutamate kinase, operably linked to suitable regulatory sequences, wherein expression of the chimeric gene results in production of levels of the encoded protein in a transformed host cell that is altered (i.e., increased or decreased) from the level produced in an untransformed host cell.

In a further embodiment, the instant invention concerns a transformed host cell comprising in its genome a chimeric gene encoding an N-acetyl glutamate kinase, operably linked to suitable regulatory sequences. Expression of the chimeric gene results in production of altered levels of the encoded protein in the transformed host cell. The transformed host cell can be of eukaryotic or prokaryotic origin, and include cells derived from higher plants and microorganisms. The invention also includes transformed plants that arise from transformed host cells of higher plants, and seeds derived from such transformed plants.

An additional embodiment of the instant invention concerns a method of altering the level of expression of an N-acetyl glutamate kinase in a transformed host cell comprising: a) transforming a host cell with a chimeric gene comprising a nucleic acid fragment encoding an N-acetyl glutamate kinase; and b) growing the transformed host cell under conditions that are suitable for expression of the chimeric gene wherein expression of the chimeric gene results in production of altered levels of N-acetyl glutamate kinase in the transformed host cell.

An addition embodiment of the instant invention concerns a method for obtaining a nucleic acid fragment encoding all or a substantial portion of an amino acid sequence encoding an N-acetyl glutamate kinase.

A further embodiment of the instant invention is a method for evaluating at least one compound for its ability to inhibit the activity of an N-acetyl glutamate kinase, the method comprising the steps of: (a) transforming a host cell with a chimeric gene comprising a nucleic acid fragment encoding an N-acetyl glutamate kinase, operably linked to suitable regulatory sequences; (b) growing the transformed host cell under conditions that are suitable for expression of the chimeric gene wherein expression of the chimeric gene results in production of N-acetyl glutamate kinase in the transformed host cell; (c) optionally purifying the N-acetyl glutamate kinase expressed by the transformed host cell; (d) treating the N-acetyl glutamate kinase with a compound to be tested; and (e) comparing the activity of the N-acetyl glutamate kinase that has been treated with a test compound to the activity of an untreated N-acetyl glutamate kinase, thereby selecting compounds with potential for inhibitory activity.

BRIEF DESCRIPTION OF THE DRAWING AND SEQUENCE DESCRIPTIONS

The invention can be more fully understood from the following detailed description and the accompanying drawing and Sequence Listing which form a part of this application.

Figure 1 depicts the amino acid sequence alignment between the N-Acetylglutamate Kinase from *Synechocystis sp.* (NCBI General Identifier No. 1652434; SEQ ID NO:11), the instant corn contig assembled of clones cr1n.pk0113.g3 and p0111.cipmf81r (SEQ ID NO:2), the instant rice clone rls72.pk0023.c6 (SEQ ID NO:4), the instant soybean clone sdp2c.pk010.h6 (SEQ ID NO:6) and the instant wheat contig assembled of clones wlk1.pk0019.d6 and wl1.pk0009.a7. The top row indicates with asterisks (*) the amino acids conserved among all sequences and with plus signs (+) the amino acids conserved only among the plant sequences. Dashes are used by the program to maximize the alignment of the sequences. The bottom row contains the consensus sequence (SEQ ID NO:12) comprising all the amino acids which are conserved among all plant sequences.

Table 1 lists the polypeptides that are described herein, the plant from which the clones originated, the designation of the cDNA clones that comprise the nucleic acid fragments encoding polypeptides representing all or a substantial portion of these polypeptides, and the corresponding identifier (SEQ ID NO:) as used in the attached Sequence Listing. The sequence descriptions and Sequence Listing attached hereto comply with the rules governing nucleotide and/or amino acid sequence disclosures in patent applications as set forth in 37 C.F.R. §1.821-1.825.

TABLE 1

N-Acetylglutamate Kinase

Plant	Clone Designation	SEQ ID NO: (Nucleotide)	SEQ ID NO: (Amino Acid)
Corn	Contig of: p0111.cipmf81r cr1n.pk0113.g3	1	2
Rice	rls72.pk0023.c6	3	4
Soybean	sdp2c.pk010.h6	5	6
Wheat	Contig of: wlk1.pk0019.d6' wl1.pk0009.a7	7	8
Wheat	wr1.pk0080.h2:fis	9	10

The Sequence Listing contains the one letter code for nucleotide sequence characters and the three letter codes for amino acids as defined in conformity with the IUPAC-IUBMB standards described in *Nucleic Acids Research 13*:3021-3030 (1985) and in the *Biochemical Journal 219 (No. 2)*:345-373 (1984) which are herein incorporated by reference. The symbols and format used for nucleotide and amino acid sequence data comply with the rules set forth in 37 C.F.R. §1.822.

DETAILED DESCRIPTION OF THE INVENTION

In the context of this disclosure, a number of terms shall be utilized. As used herein, a "nucleic acid fragment" is a polymer of RNA or DNA that is single- or double-stranded, optionally containing synthetic, non-natural or altered nucleotide bases. A nucleic acid fragment in the form of a polymer of DNA may be comprised of one or more segments of cDNA, genomic DNA or synthetic DNA.

As used herein, "contig" refers to a nucleotide sequence that is assembled from two or more constituent nucleotide sequences that share common or overlapping regions of sequence homology. For example, the nucleotide sequences of two or more nucleic acid fragments can be compared and aligned in order to identify common or overlapping sequences. Where common or overlapping sequences exist between two or more nucleic acid fragments, the sequences (and thus their corresponding nucleic acid fragments) can be assembled into a single contiguous nucleotide sequence.

As used herein, "substantially similar" refers to nucleic acid fragments wherein changes in one or more nucleotide bases results in substitution of one or more amino acids, but do not affect the functional properties of the polypeptide encoded by the nucleotide sequence. "Substantially similar" also refers to nucleic acid fragments wherein changes in one or more nucleotide bases does not affect the ability of the nucleic acid fragment to mediate alteration of gene expression by gene silencing through for example antisense or co-suppression technology. "Substantially similar" also refers to modifications of the nucleic acid fragments of the instant invention such as deletion or insertion of one or more nucleotides that do not substantially affect the functional properties of the resulting transcript vis-à-vis the ability to mediate gene silencing or alteration of the functional properties of the resulting protein molecule. It is therefore understood that the invention encompasses more than the specific exemplary nucleotide or amino acid sequences and includes functional equivalents thereof.

For example, it is well known in the art that antisense suppression and co-suppression of gene expression may be accomplished using nucleic acid fragments representing less than the entire coding region of a gene, and by nucleic acid fragments that do not share 100% sequence identity with the gene to be suppressed. Moreover, alterations in a nucleic acid fragment which result in the production of a chemically equivalent amino acid at a given site, but do not effect the functional properties of the encoded polypeptide, are well known in the art. Thus, a codon for the amino acid alanine, a hydrophobic amino acid, may be substituted by a codon encoding another less hydrophobic residue, such as glycine, or a more hydrophobic residue, such as valine, leucine, or isoleucine. Similarly, changes which result in substitution of one negatively charged residue for another, such as aspartic acid for glutamic acid, or one positively charged residue for another, such as lysine for arginine, can also be expected to produce a functionally equivalent product. Nucleotide changes which

result in alteration of the N-terminal and C-terminal portions of the polypeptide molecule would also not be expected to alter the activity of the polypeptide. Each of the proposed modifications is well within the routine skill in the art, as is determination of retention of biological activity of the encoded products.

Moreover, substantially similar nucleic acid fragments may also be characterized by their ability to hybridize. Estimates of such homology are provided by either DNA-DNA or DNA-RNA hybridization under conditions of stringency as is well understood by those skilled in the art (Hames and Higgins, Eds. (1985) Nucleic Acid Hybridisation, IRL Press, Oxford, U.K.). Stringency conditions can be adjusted to screen for moderately similar fragments, such as homologous sequences from distantly related organisms, to highly similar fragments, such as genes that duplicate functional enzymes from closely related organisms. Post-hybridization washes determine stringency conditions. One set of preferred conditions uses a series of washes starting with 6X SSC, 0.5% SDS at room temperature for 15 min, then repeated with 2X SSC, 0.5% SDS at 45°C for 30 min. and then repeated twice with 0.2X SSC, 0.5% SDS at 50°C for 30 min. A more preferred set of stringent conditions uses higher temperatures in which the washes are identical to those above except for the temperature of the final two 30 min washes in 0.2x SSC, 0.5% SDS was increased to 60°C. Another preferred set of highly stringent conditions uses two final washes in 0.1X SSC, 0.1% SDS at 65°C.

Substantially similar nucleic acid fragments of the instant invention may also be characterized by the percent identity of the amino acid sequences that they encode to the amino acid sequences disclosed herein, as determined by algorithms commonly employed by those skilled in this art. Preferred are those nucleic acid fragments whose nucleotide sequences encode amino acid sequences that are 80% identical to the amino acid sequences reported herein. More preferred nucleic acid fragments encode amino acid sequences that are 90% identical to the amino acid sequences reported herein. Most preferred are nucleic acid fragments that encode amino acid sequences that are 95% identical to the amino acid sequences reported herein. Sequence alignments and percent identity calculations were performed using the Megalign program of the LASARGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Multiple alignment of the sequences was performed using the Clustal method of alignment (Higgins and Sharp (1989) *CABIOS* 5:151-153) with the default parameters (GAP PENALTY=10, GAP LENGTH PENALTY=10). Default parameters for pairwise alignments using the Clustal method were KTUPLE 1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5.

A "substantial portion" of an amino acid or nucleotide sequence comprises an amino acid or a nucleotide sequence that is sufficient to afford putative identification of the protein or gene that the amino acid or nucleotide sequence comprises. Amino acid and nucleotide sequences can be evaluated either manually by one skilled in the art, or by using computer-

based sequence comparison and identification tools that employ algorithms such as BLAST (Basic Local Alignment Search Tool; Altschul et al. (1993) *J. Mol. Biol. 215*:403-410; see also www.ncbi.nlm.nih.gov/BLAST/). In general, a sequence of ten or more contiguous amino acids or thirty or more contiguous nucleotides is necessary in order to putatively identify a polypeptide or nucleic acid sequence as homologous to a known protein or gene. Moreover, with respect to nucleotide sequences, gene-specific oligonucleotide probes comprising 30 or more contiguous nucleotides may be used in sequence-dependent methods of gene identification (e.g., Southern hybridization) and isolation (e.g., *in situ* hybridization of bacterial colonies or bacteriophage plaques). In addition, short oligonucleotides of 12 or more nucleotides may be used as amplification primers in PCR in order to obtain a particular nucleic acid fragment comprising the primers. Accordingly, a "substantial portion" of a nucleotide sequence comprises a nucleotide sequence that will afford specific identification and/or isolation of a nucleic acid fragment comprising the sequence. The instant specification teaches amino acid and nucleotide sequences encoding polypeptides that comprise one or more particular plant proteins. The skilled artisan, having the benefit of the sequences as reported herein, may now use all or a substantial portion of the disclosed sequences for purposes known to those skilled in this art. Accordingly, the instant invention comprises the complete sequences as reported in the accompanying Sequence Listing, as well as substantial portions of those sequences as defined above.

"Codon degeneracy" refers to divergence in the genetic code permitting variation of the nucleotide sequence without effecting the amino acid sequence of an encoded polypeptide. Accordingly, the instant invention relates to any nucleic acid fragment comprising a nucleotide sequence that encodes all or a substantial portion of the amino acid sequences set forth herein. The skilled artisan is well aware of the "codon-bias" exhibited by a specific host cell in usage of nucleotide codons to specify a given amino acid. Therefore, when synthesizing a nucleic acid fragment for improved expression in a host cell, it is desirable to design the nucleic acid fragment such that its frequency of codon usage approaches the frequency of preferred codon usage of the host cell.

"Synthetic nucleic acid fragments" can be assembled from oligonucleotide building blocks that are chemically synthesized using procedures known to those skilled in the art. These building blocks are ligated and annealed to form larger nucleic acid fragments which may then be enzymatically assembled to construct the entire desired nucleic acid fragment. "Chemically synthesized", as related to nucleic acid fragment, means that the component nucleotides were assembled *in vitro*. Manual chemical synthesis of nucleic acid fragments may be accomplished using well established procedures, or automated chemical synthesis can be performed using one of a number of commercially available machines. Accordingly, the nucleic acid fragments can be tailored for optimal gene expression based on optimization of nucleotide sequence to reflect the codon bias of the host cell. The skilled artisan

appreciates the likelihood of successful gene expression if codon usage is biased towards those codons favored by the host. Determination of preferred codons can be based on a survey of genes derived from the host cell where sequence information is available.

"Gene" refers to a nucleic acid fragment that expresses a specific protein, including regulatory sequences preceding (5' non-coding sequences) and following (3' non-coding sequences) the coding sequence. "Native gene" refers to a gene as found in nature with its own regulatory sequences. "Chimeric gene" refers any gene that is not a native gene, comprising regulatory and coding sequences that are not found together in nature. Accordingly, a chimeric gene may comprise regulatory sequences and coding sequences that are derived from different sources, or regulatory sequences and coding sequences derived from the same source, but arranged in a manner different than that found in nature. "Endogenous gene" refers to a native gene in its natural location in the genome of an organism. A "foreign" gene refers to a gene not normally found in the host organism, but that is introduced into the host organism by gene transfer. Foreign genes can comprise native genes inserted into a non-native organism, or chimeric genes. A "transgene" is a gene that has been introduced into the genome by a transformation procedure.

"Coding sequence" refers to a nucleotide sequence that codes for a specific amino acid sequence. "Regulatory sequences" refer to nucleotide sequences located upstream (5' non-coding sequences), within, or downstream (3' non-coding sequences) of a coding sequence, and which influence the transcription, RNA processing or stability, or translation of the associated coding sequence. Regulatory sequences may include promoters, translation leader sequences, introns, and polyadenylation recognition sequences.

"Promoter" refers to a nucleotide sequence capable of controlling the expression of a coding sequence or functional RNA. In general, a coding sequence is located 3' to a promoter sequence. The promoter sequence consists of proximal and more distal upstream elements, the latter elements often referred to as enhancers. Accordingly, an "enhancer" is a nucleotide sequence which can stimulate promoter activity and may be an innate element of the promoter or a heterologous element inserted to enhance the level or tissue-specificity of a promoter. Promoters may be derived in their entirety from a native gene, or be composed of different elements derived from different promoters found in nature, or even comprise synthetic nucleotide segments. It is understood by those skilled in the art that different promoters may direct the expression of a gene in different tissues or cell types, or at different stages of development, or in response to different environmental conditions. Promoters which cause a nucleic acid fragment to be expressed in most cell types at most times are commonly referred to as "constitutive promoters". New promoters of various types useful in plant cells are constantly being discovered; numerous examples may be found in the compilation by Okamuro and Goldberg (1989) *Biochemistry of Plants 15*:1-82. It is further recognized that since in most cases the exact boundaries of regulatory sequences

have not been completely defined, nucleic acid fragments of different lengths may have identical promoter activity.

The "translation leader sequence" refers to a nucleotide sequence located between the promoter sequence of a gene and the coding sequence. The translation leader sequence is present in the fully processed mRNA upstream of the translation start sequence. The translation leader sequence may affect processing of the primary transcript to mRNA, mRNA stability or translation efficiency. Examples of translation leader sequences have been described (Turner and Foster (1995) *Molecular Biotechnology 3*:225).

The "3' non-coding sequences" refer to nucleotide sequences located downstream of a coding sequence and include polyadenylation recognition sequences and other sequences encoding regulatory signals capable of affecting mRNA processing or gene expression. The polyadenylation signal is usually characterized by affecting the addition of polyadenylic acid tracts to the 3' end of the mRNA precursor. The use of different 3' non-coding sequences is exemplified by Ingelbrecht et al. (1989) *Plant Cell 1*:671-680.

"RNA transcript" refers to the product resulting from RNA polymerase-catalyzed transcription of a DNA sequence. When the RNA transcript is a perfect complementary copy of the DNA sequence, it is referred to as the primary transcript or it may be a RNA sequence derived from posttranscriptional processing of the primary transcript and is referred to as the mature RNA. "Messenger RNA (mRNA)" refers to the RNA that is without introns and that can be translated into polypeptide by the cell. "cDNA" refers to a double-stranded DNA that is complementary to and derived from mRNA. "Sense" RNA refers to an RNA transcript that includes the mRNA and so can be translated into a polypeptide by the cell. "Antisense RNA" refers to an RNA transcript that is complementary to all or part of a target primary transcript or mRNA and that blocks the expression of a target gene (see U.S. Patent No. 5,107,065, incorporated herein by reference). The complementarity of an antisense RNA may be with any part of the specific nucleotide sequence, i.e., at the 5' non-coding sequence, 3' non-coding sequence, introns, or the coding sequence. "Functional RNA" refers to sense RNA, antisense RNA, ribozyme RNA, or other RNA that may not be translated but yet has an effect on cellular processes.

The term "operably linked" refers to the association of two or more nucleic acid fragments on a single nucleic acid fragment so that the function of one is affected by the other. For example, a promoter is operably linked with a coding sequence when it is capable of affecting the expression of that coding sequence (i.e., that the coding sequence is under the transcriptional control of the promoter). Coding sequences can be operably linked to regulatory sequences in sense or antisense orientation.

The term "expression", as used herein, refers to the transcription and stable accumulation of sense (mRNA) or antisense RNA derived from the nucleic acid fragment of the invention. Expression may also refer to translation of mRNA into a polypeptide.

"Antisense inhibition" refers to the production of antisense RNA transcripts capable of suppressing the expression of the target protein. "Overexpression" refers to the production of a gene product in transgenic organisms that exceeds levels of production in normal or non-transformed organisms. "Co-suppression" refers to the production of sense RNA transcripts capable of suppressing the expression of identical or substantially similar foreign or endogenous genes (U.S. Patent No. 5,231,020, incorporated herein by reference).

"Altered levels" refers to the production of gene product(s) in transgenic organisms in amounts or proportions that differ from that of normal or non-transformed organisms.

"Mature" protein refers to a post-translationally processed polypeptide; i.e., one from which any pre- or propeptides present in the primary translation product have been removed. "Precursor" protein refers to the primary product of translation of mRNA; i.e., with pre- and propeptides still present. Pre- and propeptides may be but are not limited to intracellular localization signals.

A "chloroplast transit peptide" is an amino acid sequence which is translated in conjunction with a protein and directs the protein to the chloroplast or other plastid types present in the cell in which the protein is made. "Chloroplast transit sequence" refers to a nucleotide sequence that encodes a chloroplast transit peptide. A "signal peptide" is an amino acid sequence which is translated in conjunction with a protein and directs the protein to the secretory system (Chrispeels (1991) *Ann. Rev. Plant Phys. Plant Mol. Biol. 42*:21-53). If the protein is to be directed to a vacuole, a vacuolar targeting signal (*supra*) can further be added, or if to the endoplasmic reticulum, an endoplasmic reticulum retention signal (*supra*) may be added. If the protein is to be directed to the nucleus, any signal peptide present should be removed and instead a nuclear localization signal included (Raikhel (1992) *Plant Phys. 100*:1627-1632).

"Transformation" refers to the transfer of a nucleic acid fragment into the genome of a host organism, resulting in genetically stable inheritance. Host organisms containing the transformed nucleic acid fragments are referred to as "transgenic" organisms. Examples of methods of plant transformation include *Agrobacterium*-mediated transformation (De Blaere et al. (1987) *Meth. Enzymol. 143*:277) and particle-accelerated or "gene gun" transformation technology (Klein et al. (1987) *Nature (London) 327*:70-73; U.S. Patent No. 4,945,050, incorporated herein by reference).

Standard recombinant DNA and molecular cloning techniques used herein are well known in the art and are described more fully in Sambrook et al. *Molecular Cloning: A Laboratory Manual*; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, 1989 (hereinafter "Maniatis").

Nucleic acid fragments encoding at least a portion of several N-acetylglutamate kinases have been isolated and identified by comparison of random plant cDNA sequences to public databases containing nucleotide and protein sequences using the BLAST

algorithms well known to those skilled in the art. The nucleic acid fragments of the instant invention may be used to isolate cDNAs and genes encoding homologous proteins from the same or other plant species. Isolation of homologous genes using sequence-dependent protocols is well known in the art. Examples of sequence-dependent protocols include, but are not limited to, methods of nucleic acid hybridization, and methods of DNA and RNA amplification as exemplified by various uses of nucleic acid amplification technologies (e.g., polymerase chain reaction, ligase chain reaction).

For example, genes encoding other N-acetylglutamate kinases, either as cDNAs or genomic DNAs, could be isolated directly by using all or a portion of the instant nucleic acid fragments as DNA hybridization probes to screen libraries from any desired plant employing methodology well known to those skilled in the art. Specific oligonucleotide probes based upon the instant nucleic acid sequences can be designed and synthesized by methods known in the art (Maniatis). Moreover, the entire sequences can be used directly to synthesize DNA probes by methods known to the skilled artisan such as random primer DNA labeling, nick translation, or end-labeling techniques, or RNA probes using available *in vitro* transcription systems. In addition, specific primers can be designed and used to amplify a part or all of the instant sequences. The resulting amplification products can be labeled directly during amplification reactions or labeled after amplification reactions, and used as probes to isolate full length cDNA or genomic fragments under conditions of appropriate stringency.

In addition, two short segments of the instant nucleic acid fragments may be used in polymerase chain reaction protocols to amplify longer nucleic acid fragments encoding homologous genes from DNA or RNA. The polymerase chain reaction may also be performed on a library of cloned nucleic acid fragments wherein the sequence of one primer is derived from the instant nucleic acid fragments, and the sequence of the other primer takes advantage of the presence of the polyadenylic acid tracts to the 3' end of the mRNA precursor encoding plant genes. Alternatively, the second primer sequence may be based upon sequences derived from the cloning vector. For example, the skilled artisan can follow the RACE protocol (Frohman et al. (1988) *Proc. Natl. Acad. Sci. USA 85*:8998) to generate cDNAs by using PCR to amplify copies of the region between a single point in the transcript and the 3' or 5' end. Primers oriented in the 3' and 5' directions can be designed from the instant sequences. Using commercially available 3' RACE or 5' RACE systems (BRL), specific 3' or 5' cDNA fragments can be isolated (Ohara et al. (1989) *Proc. Natl. Acad. Sci. USA 86*:5673; Loh et al. (1989) *Science 243*:217). Products generated by the 3' and 5' RACE procedures can be combined to generate full-length cDNAs (Frohman and Martin (1989) *Techniques 1*:165).

Availability of the instant nucleotide and deduced amino acid sequences facilitates immunological screening of cDNA expression libraries. Synthetic peptides representing

portions of the instant amino acid sequences may be synthesized. These peptides can be used to immunize animals to produce polyclonal or monoclonal antibodies with specificity for peptides or proteins comprising the amino acid sequences. These antibodies can be then be used to screen cDNA expression libraries to isolate full-length cDNA clones of interest (Lerner (1984) *Adv. Immunol. 36*:1; Maniatis).

The nucleic acid fragments of the instant invention may be used to create transgenic plants in which the disclosed polypeptides are present at higher or lower levels than normal or in cell types or developmental stages in which they are not normally found. This would have the effect of altering the level of arginine in those cells. Extra arginine resulting from an increase in ornithine biosynthesis may have nutraceutical utility. Also, N-acetylglutamate kinase is a potential herbicide target and may be the site of TZO-type chemistry. Overexpression of the gene coding for this enzyme, isolation of the enzyme and introduction into a screen based on its activity, may provide a means of identifying chemistry with herbicidal or fungicidal activity.

Overexpression of the proteins of the instant invention may be accomplished by first constructing a chimeric gene in which the coding region is operably linked to a promoter capable of directing expression of a gene in the desired tissues at the desired stage of development. For reasons of convenience, the chimeric gene may comprise promoter sequences and translation leader sequences derived from the same genes. 3' Non-coding sequences encoding transcription termination signals may also be provided. The instant chimeric gene may also comprise one or more introns in order to facilitate gene expression.

Plasmid vectors comprising the instant chimeric gene can then constructed. The choice of plasmid vector is dependent upon the method that will be used to transform host plants. The skilled artisan is well aware of the genetic elements that must be present on the plasmid vector in order to successfully transform, select and propagate host cells containing the chimeric gene. The skilled artisan will also recognize that different independent transformation events will result in different levels and patterns of expression (Jones et al. (1985) *EMBO J. 4*:2411-2418; De Almeida et al. (1989) *Mol. Gen. Genetics 218*:78-86), and thus that multiple events must be screened in order to obtain lines displaying the desired expression level and pattern. Such screening may be accomplished by Southern analysis of DNA, Northern analysis of mRNA expression, Western analysis of protein expression, or phenotypic analysis.

For some applications it may be useful to direct the instant polypeptide to different cellular compartments, or to facilitate its secretion from the cell. It is thus envisioned that the chimeric gene described above may be further supplemented by altering the coding sequence to encode the instant polypeptide with appropriate intracellular targeting sequences such as transit sequences (Keegstra (1989) *Cell 56*:247-253), signal sequences or sequences encoding endoplasmic reticulum localization (Chrispeels (1991) *Ann. Rev. Plant Phys. Plant*

Mol. Biol. 42:21-53), or nuclear localization signals (Raikhel (1992) *Plant Phys.100*:1627-1632) added and/or with targeting sequences that are already present removed. While the references cited give examples of each of these, the list is not exhaustive and more targeting signals of utility may be discovered in the future.

It may also be desirable to reduce or eliminate expression of genes encoding the instant polypeptide in plants for some applications. In order to accomplish this, a chimeric gene designed for co-suppression of the instant polypeptide can be constructed by linking a gene or gene fragment encoding that polypeptide to plant promoter sequences. Alternatively, a chimeric gene designed to express antisense RNA for all or part of the instant nucleic acid fragment can be constructed by linking the gene or gene fragment in reverse orientation to plant promoter sequences. Either the co-suppression or antisense chimeric genes could be introduced into plants via transformation wherein expression of the corresponding endogenous genes are reduced or eliminated.

Molecular genetic solutions to the generation of plants with altered gene expression have a decided advantage over more traditional plant breeding approaches. Changes in plant phenotypes can be produced by specifically inhibiting expression of one or more genes by antisense inhibition or cosuppression (U. S. Patent Nos. 5,190,931, 5,107,065 and 5,283,323). An antisense or cosuppression construct would act as a dominant negative regulator of gene activity. While conventional mutations can yield negative regulation of gene activity these effects are most likely recessive. The dominant negative regulation available with a transgenic approach may be advantageous from a breeding perspective. In addition, the ability to restrict the expression of specific phenotype to the reproductive tissues of the plant by the use of tissue specific promoters may confer agronomic advantages relative to conventional mutations which may have an effect in all tissues in which a mutant gene is ordinarily expressed.

The person skilled in the art will know that special considerations are associated with the use of antisense or cosuppresion technologies in order to reduce expression of particular genes. For example, the proper level of expression of sense or antisense genes may require the use of different chimeric genes utilizing different regulatory elements known to the skilled artisan. Once transgenic plants are obtained by one of the methods described above, it will be necessary to screen individual transgenics for those that most effectively display the desired phenotype. Accordingly, the skilled artisan will develop methods for screening large numbers of transformants. The nature of these screens will generally be chosen on practical grounds, and is not an inherent part of the invention. For example, one can screen by looking for changes in gene expression by using antibodies specific for the protein encoded by the gene being suppressed, or one could establish assays that specifically measure enzyme activity. A preferred method will be one which allows large numbers of

samples to be processed rapidly, since it will be expected that a large number of transformants will be negative for the desired phenotype.

The instant polypeptide (or portions thereof) may be produced in heterologous host cells, particularly in the cells of microbial hosts, and can be used to prepare antibodies to the these proteins by methods well known to those skilled in the art. The antibodies are useful for detecting the polypeptide of the instant invention *in situ* in cells or *in vitro* in cell extracts. Preferred heterologous host cells for production of the instant polypeptide are microbial hosts. Microbial expression systems and expression vectors containing regulatory sequences that direct high level expression of foreign proteins are well known to those skilled in the art. Any of these could be used to construct a chimeric gene for production of the instant polypeptide. This chimeric gene could then be introduced into appropriate microorganisms via transformation to provide high level expression of the encoded N-acetylglutamate kinase. An example of a vector for high level expression of the instant polypeptide in a bacterial host is provided (Example 6).

Additionally, the instant polypeptide can be used as a target to facilitate design and/or identification of inhibitors of those enzymes that may be useful as herbicides. This is desirable because the polypeptide described herein catalyzes a step in ornithine biosynthesis. Accordingly, inhibition of the activity of the enzyme described herein could lead to inhibition plant growth. Thus, the instant polypeptide could be appropriate for new herbicide discovery and design.

All or a substantial portion of the nucleic acid fragments of the instant invention may also be used as probes for genetically and physically mapping the genes that they are a part of, and as markers for traits linked to those genes. Such information may be useful in plant breeding in order to develop lines with desired phenotypes. For example, the instant nucleic acid fragments may be used as restriction fragment length polymorphism (RFLP) markers. Southern blots (Maniatis) of restriction-digested plant genomic DNA may be probed with the nucleic acid fragments of the instant invention. The resulting banding patterns may then be subjected to genetic analyses using computer programs such as MapMaker (Lander et al. (1987) *Genomics 1*:174-181) in order to construct a genetic map. In addition, the nucleic acid fragments of the instant invention may be used to probe Southern blots containing restriction endonuclease-treated genomic DNAs of a set of individuals representing parent and progeny of a defined genetic cross. Segregation of the DNA polymorphisms is noted and used to calculate the position of the instant nucleic acid sequence in the genetic map previously obtained using this population (Botstein et al. (1980) *Am. J. Hum. Genet. 32*:314-331).

The production and use of plant gene-derived probes for use in genetic mapping is described in Bernatzky and Tanksley (1986) *Plant Mol. Biol. Reporter 4(1)*:37-41. Numerous publications describe genetic mapping of specific cDNA clones using the

methodology outlined above or variations thereof. For example, F2 intercross populations, backcross populations, randomly mated populations, near isogenic lines, and other sets of individuals may be used for mapping. Such methodologies are well known to those skilled in the art.

Nucleic acid probes derived from the instant nucleic acid sequences may also be used for physical mapping (i.e., placement of sequences on physical maps: *see* Hoheisel et al. In: *Nonmammalian Genomic Analysis: A Practical Guide*. Academic press 1996, pp. 319-346, and references cited therein).

In another embodiment, nucleic acid probes derived from the instant nucleic acid sequences may be used in direct fluorescence *in situ* hybridization (FISH) mapping (Trask (1991) *Trends Genet. 7*:149-154). Although current methods of FISH mapping favor use of large clones (several to several hundred KB; see Laan et al. (1995) *Genome Research 5*:13-20), improvements in sensitivity may allow performance of FISH mapping using shorter probes.

A variety of nucleic acid amplification-based methods of genetic and physical mapping may be carried out using the instant nucleic acid sequences. Examples include allele-specific amplification (Kazazian (1989) *J. Lab. Clin. Med. 114(2)*:95-96), polymorphism of PCR-amplified fragments (CAPS; Sheffield et al. (1993) *Genomics 16*:325-332), allele-specific ligation (Landegren et al. (1988) *Science 241*:1077-1080), nucleotide extension reactions (Sokolov (1990) *Nucleic Acid Res. 18*:3671), Radiation Hybrid Mapping (Walter et al. (1997) *Nature Genetics 7*:22-28) and Happy Mapping (Dear and Cook (1989) *Nucleic Acid Res. 17*:6795-6807). For these methods, the sequence of a nucleic acid fragment is used to design and produce primer pairs for use in the amplification reaction or in primer extension reactions. The design of such primers is well known to those skilled in the art. In methods employing PCR-based genetic mapping, it may be necessary to identify DNA sequence differences between the parents of the mapping cross in the region corresponding to the instant nucleic acid sequence. This, however, is generally not necessary for mapping methods.

Loss of function mutant phenotypes may be identified for the instant cDNA clones either by targeted gene disruption protocols or by identifying specific mutants for these genes contained in a maize population carrying mutations in all possible genes (Ballinger and Benzer (1989) *Proc. Natl. Acad. Sci USA 86*:9402; Koes et al. (1995) *Proc. Natl. Acad. Sci USA 92*:8149; Bensen et al. (1995) *Plant Cell 7*:75). The latter approach may be accomplished in two ways. First, short segments of the instant nucleic acid fragments may be used in polymerase chain reaction protocols in conjunction with a mutation tag sequence primer on DNAs prepared from a population of plants in which Mutator transposons or some other mutation-causing DNA element has been introduced (see Bensen, *supra*). The amplification of a specific DNA fragment with these primers indicates the insertion of the

mutation tag element in or near the plant gene encoding the instant polypeptide. Alternatively, the instant nucleic acid fragment may be used as a hybridization probe against PCR amplification products generated from the mutation population using the mutation tag sequence primer in conjunction with an arbitrary genomic site primer. such as that for a restriction enzyme site-anchored synthetic adaptor. With either method, a plant containing a mutation in the endogenous gene encoding the instant polypeptide can be identified and obtained. This mutant plant can then be used to determine or confirm the natural function of the instant polypeptide disclosed herein.

EXAMPLES

The present invention is further defined in the following Examples, in which all parts and percentages are by weight and degrees are Celsius, unless otherwise stated. It should be understood that these Examples, while indicating preferred embodiments of the invention, are given by way of illustration only. From the above discussion and these Examples, one skilled in the art can ascertain the essential characteristics of this invention, and without departing from the spirit and scope thereof, can make various changes and modifications of the invention to adapt it to various usages and conditions.

EXAMPLE 1

Composition of cDNA Libraries: Isolation and Sequencing of cDNA Clones

cDNA libraries representing mRNAs from various corn, rice, soybean and wheat tissues were prepared. The characteristics of the libraries are described below.

TABLE 2

cDNA Libraries from Corn, Rice, Soybean and Wheat

Library	Tissue	Clone
cr1n	Corn Root From 7 Day Old Seedlings*	cr1n.pk0113.g3
p0111	Corn V6 Stage Leaf Tissue Minus Midrib + UV, screened 1 Pool of PR + UV 3h; PR +UV 24 h; PR+UV 48h; and PR+UV 7 days	p0111.cipmf81r
rls72	Rice Leaf 15 Days After Germination, 72 Hours After Infection of Strain *Magaporthe grisea* 4360-R-67 (AVR2-YAMO); Susceptible	rls72.pk0023.c6
sdp2c	Soybean Developing Pods (6-7 mm)	sdp2c.pk010.h6
wl1	Wheat Leaf From 7 Day Old Seedling	wl1.pk0009.a7
wlk1	Wheat Seedlings 1 Hour After Treatment With Herbicide**	wlk1.pk0019.d6
wr1	Wheat Root From 7 Day Old Seedling	wr1.pk0080.h2

*These libraries were normalized essentially as described in U.S. Patent No. 5,482,845, incorporated herein by reference.

**Application of 6-iodo-2-propoxy-3-propyl-4(3*H*)-quinazolinone; synthesis and methods of using this compound are described in USSN 08/545,827, incorporated herein by reference.

cDNA libraries may be prepared by any one of many methods available. For example, the cDNAs may be introduced into plasmid vectors by first preparing the cDNA libraries in Uni-ZAP™ XR vectors according to the manufacturer's protocol (Stratagene Cloning Systems, La Jolla, CA). The Uni-ZAP™ XR libraries are converted into plasmid libraries according to the protocol provided by Stratagene. Upon conversion, cDNA inserts will be contained in the plasmid vector pBluescript. In addition, the cDNAs may be introduced directly into precut Bluescript II SK(+) vectors (Stratagene) using T4 DNA ligase (New England Biolabs), followed by transfection into DH10B cells according to the manufacturer's protocol (GIBCO BRL Products). Once the cDNA inserts are in plasmid vectors, plasmid DNAs are prepared from randomly picked bacterial colonies containing recombinant pBluescript plasmids, or the insert cDNA sequences are amplified via polymerase chain reaction using primers specific for vector sequences flanking the inserted cDNA sequences. Amplified insert DNAs or plasmid DNAs are sequenced in dye-primer sequencing reactions to generate partial cDNA sequences (expressed sequence tags or "ESTs"; see Adams et al., (1991) *Science 252*:1651). The resulting ESTs are analyzed using a Perkin Elmer Model 377 fluorescent sequencer.

EXAMPLE 2

Identification of cDNA Clones

cDNA clones encoding N-acetylglutamate kinase were identified by conducting BLAST (Basic Local Alignment Search Tool; Altschul et al. (1993) *J. Mol. Biol. 215*:403-410; see also www.ncbi.nlm.nih.gov/BLAST/) searches for similarity to sequences contained in the BLAST "nr" database (comprising all non-redundant GenBank CDS translations, sequences derived from the 3-dimensional structure Brookhaven Protein Data Bank, the last major release of the SWISS-PROT protein sequence database, EMBL, and DDBJ databases). The cDNA sequences obtained in Example 1 were analyzed for similarity to all publicly available DNA sequences contained in the "nr" database using the BLASTN algorithm provided by the National Center for Biotechnology Information (NCBI). The DNA sequences were translated in all reading frames and compared for similarity to all publicly available protein sequences contained in the "nr" database using the BLASTX algorithm (Gish and States (1993) *Nature Genetics 3*:266-272) provided by the NCBI. For convenience, the P-value (probability) of observing a match of a cDNA sequence to a sequence contained in the searched databases merely by chance as calculated by BLAST are reported herein as "pLog" values, which represent the negative of the logarithm of the reported P-value. Accordingly, the greater the pLog value, the greater the likelihood that the cDNA sequence and the BLAST "hit" represent homologous proteins.

EXAMPLE 3

Characterization of cDNA Clones Encoding N-Acetylglutamate Kinase

The BLASTX search using the EST sequences from clones listed in Table 3 revealed similarity of the polypeptides encoded by the cDNAs to N-acetylglutamate kinase from *Synechocystis sp.* (NCBI General Identifier No. 1652434). Shown in Table 3 are the BLAST results for individual ESTs ("EST"), the sequences of the entire cDNA inserts comprising the indicated cDNA clones ("FIS"), or contigs assembled from two or more ESTs or assembled of an FIS and an EST ("Contig"):

TABLE 3

BLAST Results for Sequences Encoding Polypeptides Homologous to N-Acetylglutamate Kinase

Clone	Status	BLAST pLog Score 1652434
Contig of: p0111.cipmf81r cr1n.pk0113.g3	Contig	102.0
rls72.pk0023.c6	FIS	94.1
sdp2c.pk010.h6	FIS	95.0
Contig of: wlk1.pk0019.d6 wl1.pk0009.a7	Contig	86.0
wr1.pk0080.h2	FIS	29.0

The nucleotide sequence of wheat clone wlk1.pk0019.d6 was missing 51 nucleotides. To complete the entire sequence of a wheat N-acetylglutamate kinase 51 nucleotides were added as unknowns (n) and converted into 17 unknown amino acids (X, amino acids 144 to 160). Figure 1 presents an alignment of the amino acid sequences set forth in SEQ ID NOs:2, 4, 6 and 8 and the *Synechocystis sp.* sequence (SEQ ID NO:11). The data in Table 4 represents a calculation of the percent identity of the amino acid sequences set forth in SEQ ID NOs:2, 4, 6, 8 and 10 and the *Synechocystis sp.* sequence (SEQ ID NO:11). The amino acid sequence set forth in SEQ ID NO:10 is 89% identical over the 98 amino acid region or overlap with the amino acid sequence set forth in SEQ ID NO:8.

TABLE 4

Percent Identity of Amino Acid Sequences Deduced From the Nucleotide Sequences of cDNA Clones Encoding Polypeptides Homologous to N-Acetylglutamate Kinase

SEQ ID NO.	Percent Identity to 1652434
2	60.6
4	56.6
6	57.6
8	52.5
10	59.4

Sequence alignments and percent identity calculations were performed using the Megalign program of the LASARGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Multiple alignment of the sequences was performed using the Clustal method of alignment (Higgins and Sharp (1989) *CABIOS.* 5:151-153) with the default parameters (GAP PENALTY=10, GAP LENGTH PENALTY=10). Default parameters for pairwise alignments using the Clustal method were KTUPLE 1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5.

A "consensus sequence" was assembled using all the amino acids conserved among the plant (corn, rice, soybean and wheat) sequences (SEQ ID NO:12). The amino acids vary among the different plants creating positions where more than one amino acid may be present. Table 5 presents a list of the positions where these amino acids vary and the possible amino acids present at these positions.

TABLE 5

Amino Acid Variability Among Plant N-Acetylglutamate Kinase

Amino Acid Position in Consensus Sequence	Possible Amino Acids
2	Leu, Met
3	Leu, Ala
4, 176	Thr, Ala, Gly
5	Pro, none
6	Trp, His, none
7	Leu, none
8, 9, 46	Ser, Ala, none
10	Ser, Lys, Pro
11	Lys, Ser, Thr, Ala

Amino Acid Position in Consensus Sequence	Possible Amino Acids
12	Leu, Phe
13	Pro, Thr. none
14	Val, Leu. Asn
15	Pro, Leu
16, 182	Ser, Cys
17, 36	Pro, Thr. Ala
18	Pro, Arg. Ser
19	Pro, Val. Phe, Leu
20, 30	Ser, Lys, Thr, Arg
21	Gly, Phe. Asn
23, 50	Thr, Ala. Asn
24	Leu, Pro. Lys
25	Ser, Gly, Pro, Lys
26	Ser, Pro, Gln, Ala
27	Asn, Ala
28	His, Gln, Arg
29	Ala, Leu, Val
31	Pro, Thr
32	Leu, Ile, Ser
33	Ala, His
34, 49, 51, 52, 74. 194, 252, 265, 309	Ala, Ser
35	Pro, Ser. Phe
37	Cys, Ala. Ser, Pro
38	Arg, Pro. Thr, His
39	Arg, Gly. none
40, 43	Arg, Leu
41	Ser, Arg
42	Arg, Cys, His, Gly
44	Arg, Ala
45, 253	Ile, Leu. Val
47	Ala, Val. none
48	Thr, Val. none
53	Pro, Gln. Ser
54, 57	Ser, Ala, Pro
55	Pro, Ala. Leu

Amino Acid Position in Consensus Sequence	Possible Amino Acids
56	Ser, Leu. Ala
59	Ala, Glu. Gln
60, 227, 428	Ala, Thr
62	Thr, Ser, none
63, 191, 197, 269, 291	Glu, Ala
64	Ala, Ser. Gly
65	Leu, Gln
66	Ser, Tyr, Asn
80, 82	Arg, Lys
86, 137, 154, 179, 226, 281	Val, Ile
106	Arg, Asn
117. 215; 308	Arg, His
133	Leu, Gly, Gln
135, 209	Val, Lys
136	Gly, Asn
138	Glu, Pro
139	Gln, Val
142. 210, 272	Asp, Asn
150	Leu, Asp
151, 190	Thr, Asn
156	Glu, Ser
161, 201	Gly, Arg
166	Asn, Glu. Thr, Gln
172	Asn, Lys. Arg
173	Ile, Leu. Lys, Pro
174. 211, 237	Ala, Pro
177. 228, 234	Thr, Ser
180, 221	Gly, Ser
183	Gly, Trp, Arg
184	Lys, Met
185, 292	Asp, Glu
186. 230, 296	Gly, Ala
189, 217. 249, 259, 261	Ile, Leu
196	Asn, Lys, Asp
198, 284	Lys, Ala

Amino Acid Position in Consensus Sequence	Possible Amino Acids
199	Ala, Gly, Asp
202	Phe, Tyr
205	Glu, Gly
207	Ser, Trp, Ala, Thr
212, 220	Thr, Ala, Ser
216	Pro, Ser
219	Ala, Asp
245	Ala, Val
275	Gly, Asp
280	Val, Glu, Lys
288	Lys, Gln, Arg
290	Val, Met
294	Lys, Gln
304	Glu, Asn, Gly
405	His, Ile

Sequence alignments and BLAST scores and probabilities indicate that the nucleic acid fragments comprising the instant cDNA clones encode an almost entire wheat N-acetylglutamate kinase, entire corn, rice and soybean N-acetylglutamate kinases, and a substantial portion of a wheat N-acetylglutamate kinase variant. These sequences represent the first plant sequences encoding N-acetylglutamate kinase.

EXAMPLE 4

Expression of Chimeric Genes in Monocot Cells

A chimeric gene comprising a cDNA encoding the instant polypeptide in sense orientation with respect to the maize 27 kD zein promoter that is located 5' to the cDNA fragment, and the 10 kD zein 3' end that is located 3' to the cDNA fragment, can be constructed. The cDNA fragment of this gene may be generated by polymerase chain reaction (PCR) of the cDNA clone using appropriate oligonucleotide primers. Cloning sites (NcoI or SmaI) can be incorporated into the oligonucleotides to provide proper orientation of the DNA fragment when inserted into the digested vector pML103 as described below. Amplification is then performed in a standard PCR. The amplified DNA is then digested with restriction enzymes NcoI and SmaI and fractionated on an agarose gel. The appropriate band can be isolated from the gel and combined with a 4.9 kb NcoI-SmaI fragment of the plasmid pML103. Plasmid pML103 has been deposited under the terms of the Budapest Treaty at ATCC (American Type Culture Collection, 10801 University Blvd., Manassas,

VA 20110-2209), and bears accession number ATCC 97366. The DNA segment from pML103 contains a 1.05 kb SalI-NcoI promoter fragment of the maize 27 kD zein gene and a 0.96 kb SmaI-SalI fragment from the 3' end of the maize 10 kD zein gene in the vector pGem9Zf(+) (Promega). Vector and insert DNA can be ligated at 15°C overnight, essentially as described (Maniatis). The ligated DNA may then be used to transform *E. coli* XL1-Blue (Epicurian Coli XL-1 Blue™; Stratagene). Bacterial transformants can be screened by restriction enzyme digestion of plasmid DNA and limited nucleotide sequence analysis using the dideoxy chain termination method (Sequenase™ DNA Sequencing Kit; U.S. Biochemical). The resulting plasmid construct would comprise a chimeric gene encoding, in the 5' to 3' direction, the maize 27 kD zein promoter, a cDNA fragment encoding the instant polypeptide, and the 10 kD zein 3' region.

The chimeric gene described above can then be introduced into corn cells by the following procedure. Immature corn embryos can be dissected from developing caryopses derived from crosses of the inbred corn lines H99 and LH132. The embryos are isolated 10 to 11 days after pollination when they are 1.0 to 1.5 mm long. The embryos are then placed with the axis-side facing down and in contact with agarose-solidified N6 medium (Chu et al. (1975) *Sci. Sin. Peking* 18:659-668). The embryos are kept in the dark at 27°C. Friable embryogenic callus consisting of undifferentiated masses of cells with somatic proembryoids and embryoids borne on suspensor structures proliferates from the scutellum of these immature embryos. The embryogenic callus isolated from the primary explant can be cultured on N6 medium and sub-cultured on this medium every 2 to 3 weeks.

The plasmid, p35S/Ac (obtained from Dr. Peter Eckes, Hoechst Ag, Frankfurt, Germany) may be used in transformation experiments in order to provide for a selectable marker. This plasmid contains the *Pat* gene (see European Patent Publication 0 242 236) which encodes phosphinothricin acetyl transferase (PAT). The enzyme PAT confers resistance to herbicidal glutamine synthetase inhibitors such as phosphinothricin. The *pat* gene in p35S/Ac is under the control of the 35S promoter from Cauliflower Mosaic Virus (Odell et al. (1985) *Nature* 313:810-812) and the 3' region of the nopaline synthase gene from the T-DNA of the Ti plasmid of *Agrobacterium tumefaciens*.

The particle bombardment method (Klein et al. (1987) *Nature* 327:70-73) may be used to transfer genes to the callus culture cells. According to this method, gold particles (1 μm in diameter) are coated with DNA using the following technique. Ten μg of plasmid DNAs are added to 50 μL of a suspension of gold particles (60 mg per mL). Calcium chloride (50 μL of a 2.5 M solution) and spermidine free base (20 μL of a 1.0 M solution) are added to the particles. The suspension is vortexed during the addition of these solutions. After 10 minutes, the tubes are briefly centrifuged (5 sec at 15,000 rpm) and the supernatant removed. The particles are resuspended in 200 μL of absolute ethanol, centrifuged again and the supernatant removed. The ethanol rinse is performed again and the particles

resuspended in a final volume of 30 μL of ethanol. An aliquot (5 μL) of the DNA-coated gold particles can be placed in the center of a Kapton™ flying disc (Bio-Rad Labs). The particles are then accelerated into the corn tissue with a Biolistic™ PDS-1000/He (Bio-Rad Instruments, Hercules CA), using a helium pressure of 1000 psi, a gap distance of 0.5 cm and a flying distance of 1.0 cm.

For bombardment, the embryogenic tissue is placed on filter paper over agarose-solidified N6 medium. The tissue is arranged as a thin lawn and covered a circular area of about 5 cm in diameter. The petri dish containing the tissue can be placed in the chamber of the PDS-1000/He approximately 8 cm from the stopping screen. The air in the chamber is then evacuated to a vacuum of 28 inches of Hg. The macrocarrier is accelerated with a helium shock wave using a rupture membrane that bursts when the He pressure in the shock tube reaches 1000 psi.

Seven days after bombardment the tissue can be transferred to N6 medium that contains gluphosinate (2 mg per liter) and lacks casein or proline. The tissue continues to grow slowly on this medium. After an additional 2 weeks the tissue can be transferred to fresh N6 medium containing gluphosinate. After 6 weeks, areas of about 1 cm in diameter of actively growing callus can be identified on some of the plates containing the glufosinate-supplemented medium. These calli may continue to grow when sub-cultured on the selective medium.

Plants can be regenerated from the transgenic callus by first transferring clusters of tissue to N6 medium supplemented with 0.2 mg per liter of 2,4-D. After two weeks the tissue can be transferred to regeneration medium (Fromm et al. (1990) *Bio/Technology 8*:833-839).

EXAMPLE 5

Expression of Chimeric Genes in Dicot Cells

A seed-specific expression cassette composed of the promoter and transcription terminator from the gene encoding the β subunit of the seed storage protein phaseolin from the bean *Phaseolus vulgaris* (Doyle et al. (1986) *J. Biol. Chem. 261*:9228-9238) can be used for expression of the instant polypeptide in transformed soybean. The phaseolin cassette includes about 500 nucleotides upstream (5') from the translation initiation codon and about 1650 nucleotides downstream (3') from the translation stop codon of phaseolin. Between the 5' and 3' regions are the unique restriction endonuclease sites Nco I (which includes the ATG translation initiation codon), Sma I, Kpn I and Xba I. The entire cassette is flanked by Hind III sites.

The cDNA fragment of this gene may be generated by polymerase chain reaction (PCR) of the cDNA clone using appropriate oligonucleotide primers. Cloning sites can be incorporated into the oligonucleotides to provide proper orientation of the DNA fragment when inserted into the expression vector. Amplification is then performed as described

above, and the isolated fragment is inserted into a pUC18 vector carrying the seed expression cassette.

Soybean embroys may then be transformed with the expression vector comprising sequences encoding the instant polypeptide. To induce somatic embryos, cotyledons, 3-5 mm in length dissected from surface sterilized, immature seeds of the soybean cultivar A2872, can be cultured in the light or dark at 26°C on an appropriate agar medium for 6-10 weeks. Somatic embryos which produce secondary embryos are then excised and placed into a suitable liquid medium. After repeated selection for clusters of somatic embryos which multiplied as early, globular staged embryos, the suspensions are maintained as described below.

Soybean embryogenic suspension cultures can maintained in 35 mL liquid media on a rotary shaker, 150 rpm, at 26°C with florescent lights on a 16:8 hour day/night schedule. Cultures are subcultured every two weeks by inoculating approximately 35 mg of tissue into 35 mL of liquid medium.

Soybean embryogenic suspension cultures may then be transformed by the method of particle gun bombardment (Klein et al. (1987) *Nature* (London) *327*:70, U.S. Patent No. 4,945,050). A DuPont Biolistic™ PDS1000/HE instrument (helium retrofit) can be used for these transformations.

A selectable marker gene which can be used to facilitate soybean transformation is a chimeric gene composed of the 35S promoter from Cauliflower Mosaic Virus (Odell et al. (1985) *Nature 313*:810-812), the hygromycin phosphotransferase gene from plasmid pJR225 (from *E. coli*; Gritz et al.(1983) *Gene 25*:179-188) and the 3' region of the nopaline synthase gene from the T-DNA of the Ti plasmid of *Agrobacterium tumefaciens*. The seed expression cassette comprising the phaseolin 5' region, the fragment encoding the instant polypeptide and the phaseolin 3' region can be isolated as a restriction fragment. This fragment can then be inserted into a unique restriction site of the vector carrying the marker gene.

To 50 µL of a 60 mg/mL 1 µm gold particle suspension is added (in order): 5 µL DNA (1 µg/µL), 20 µl spermidine (0.1 M), and 50 µL $CaCl_2$ (2.5 M). The particle preparation is then agitated for three minutes, spun in a microfuge for 10 seconds and the supernatant removed. The DNA-coated particles are then washed once in 400 µL 70% ethanol and resuspended in 40 µL of anhydrous ethanol. The DNA/particle suspension can be sonicated three times for one second each. Five µL of the DNA-coated gold particles are then loaded on each macro carrier disk.

Approximately 300-400 mg of a two-week-old suspension culture is placed in an empty 60x15 mm petri dish and the residual liquid removed from the tissue with a pipette. For each transformation experiment, approximately 5-10 plates of tissue are normally bombarded. Membrane rupture pressure is set at 1100 psi and the chamber is evacuated to a vacuum of 28 inches mercury. The tissue is placed approximately 3.5 inches away from the

retaining screen and bombarded three times. Following bombardment, the tissue can be divided in half and placed back into liquid and cultured as described above.

Five to seven days post bombardment, the liquid media may be exchanged with fresh media, and eleven to twelve days post bombardment with fresh media containing 50 mg/mL hygromycin. This selective media can be refreshed weekly. Seven to eight weeks post bombardment, green, transformed tissue may be observed growing from untransformed, necrotic embryogenic clusters. Isolated green tissue is removed and inoculated into individual flasks to generate new, clonally propagated, transformed embryogenic suspension cultures. Each new line may be treated as an independent transformation event. These suspensions can then be subcultured and maintained as clusters of immature embryos or regenerated into whole plants by maturation and germination of individual somatic embryos.

EXAMPLE 6

Expression of Chimeric Genes in Microbial Cells

The cDNAs encoding the instant polypeptide can be inserted into the T7 *E. coli* expression vector pBT430. This vector is a derivative of pET-3a (Rosenberg et al. (1987) *Gene 56*:125-135) which employs the bacteriophage T7 RNA polymerase/T7 promoter system. Plasmid pBT430 was constructed by first destroying the EcoR I and Hind III sites in pET-3a at their original positions. An oligonucleotide adaptor containing EcoR I and Hind III sites was inserted at the BamH I site of pET-3a. This created pET-3aM with additional unique cloning sites for insertion of genes into the expression vector. Then, the Nde I site at the position of translation initiation was converted to an Nco I site using oligonucleotide-directed mutagenesis. The DNA sequence of pET-3aM in this region, 5'-CATATGG, was converted to 5'-CCCATGG in pBT430.

Plasmid DNA containing a cDNA may be appropriately digested to release a nucleic acid fragment encoding the protein. This fragment may then be purified on a 1% NuSieve GTG™ low melting agarose gel (FMC). Buffer and agarose contain 10 μg/ml ethidium bromide for visualization of the DNA fragment. The fragment can then be purified from the agarose gel by digestion with GELase™ (Epicentre Technologies) according to the manufacturer's instructions, ethanol precipitated, dried and resuspended in 20 μL of water. Appropriate oligonucleotide adapters may be ligated to the fragment using T4 DNA ligase (New England Biolabs, Beverly, MA). The fragment containing the ligated adapters can be purified from the excess adapters using low melting agarose as described above. The vector pBT430 is digested, dephosphorylated with alkaline phosphatase (NEB) and deproteinized with phenol/chloroform as described above. The prepared vector pBT430 and fragment can then be ligated at 16°C for 15 hours followed by transformation into DH5 electrocompetent cells (GIBCO BRL). Transformants can be selected on agar plates containing LB media and 100 μg/mL ampicillin. Transformants containing the gene encoding the instant polypeptide

are then screened for the correct orientation with respect to the T7 promoter by restriction enzyme analysis.

For high level expression, a plasmid clone with the cDNA insert in the correct orientation relative to the T7 promoter can be transformed into *E. coli* strain BL21(DE3) (Studier et al. (1986) *J. Mol. Biol. 189*:113-130). Cultures are grown in LB medium containing ampicillin (100 mg/L) at 25°C. At an optical density at 600 nm of approximately 1, IPTG (isopropylthio-β-galactoside, the inducer) can be added to a final concentration of 0.4 mM and incubation can be continued for 3 h at 25°. Cells are then harvested by centrifugation and re-suspended in 50 µL of 50 mM Tris-HCl at pH 8.0 containing 0.1 mM DTT and 0.2 mM phenyl methylsulfonyl fluoride. A small amount of 1 mm glass beads can be added and the mixture sonicated 3 times for about 5 seconds each time with a microprobe sonicator. The mixture is centrifuged and the protein concentration of the supernatant determined. One µg of protein from the soluble fraction of the culture can be separated by SDS-polyacrylamide gel electrophoresis. Gels can be observed for protein bands migrating at the expected molecular weight.

EXAMPLE 7

Evaluating Compounds for Their Ability to Inhibit the Activity of N-Acetylglutamate Kinase

The polypeptide described herein may be produced using any number of methods known to those skilled in the art. Such methods include, but are not limited to, expression in bacteria as described in Example 6, or expression in eukaryotic cell culture, *in planta*, and using viral expression systems in suitably infected organisms or cell lines. The instant polypeptide may be expressed either as mature forms of the proteins as observed *in vivo* or as fusion proteins by covalent attachment to a variety of enzymes, proteins or affinity tags. Common fusion protein partners include glutathione S-transferase ("GST"), thioredoxin ("Trx"), maltose binding protein, and C- and/or N-terminal hexahistidine polypeptide ("$(His)_6$"). The fusion proteins may be engineered with a protease recognition site at the fusion point so that fusion partners can be separated by protease digestion to yield intact mature enzyme. Examples of such proteases include thrombin, enterokinase and factor Xa. However, any protease can be used which specifically cleaves the peptide connecting the fusion protein and the enzyme.

Purification of the instant polypeptide, if desired, may utilize any number of separation technologies familiar to those skilled in the art of protein purification. Examples of such methods include, but are not limited to, homogenization, filtration, centrifugation, heat denaturation, ammonium sulfate precipitation, desalting, pH precipitation, ion exchange chromatography, hydrophobic interaction chromatography and affinity chromatography, wherein the affinity ligand represents a substrate, substrate analog or inhibitor. When the instant polypeptide are expressed as fusion proteins, the purification protocol may include

the use of an affinity resin which is specific for the fusion protein tag attached to the expressed enzyme or an affinity resin containing ligands which are specific for the enzyme. For example, the instant polypeptide may be expressed as a fusion protein coupled to the C-terminus of thioredoxin. In addition, a $(His)_6$ peptide may be engineered into the N-terminus of the fused thioredoxin moiety to afford additional opportunities for affinity purification. Other suitable affinity resins could be synthesized by linking the appropriate ligands to any suitable resin such as Sepharose-4B. In an alternate embodiment, a thioredoxin fusion protein may be eluted using dithiothreitol; however, elution may be accomplished using other reagents which interact to displace the thioredoxin from the resin. These reagents include β–mercaptoethanol or other reduced thiol. The eluted fusion protein may be subjected to further purification by traditional means as stated above, if desired. Proteolytic cleavage of the thioredoxin fusion protein and the enzyme may be accomplished after the fusion protein is purified or while the protein is still bound to the ThioBond™ affinity resin or other resin.

Crude, partially purified or purified enzyme, either alone or as a fusion protein, may be utilized in assays for the evaluation of compounds for their ability to inhibit enzymatic activation of the instant polypeptide disclosed herein. Assays may be conducted under well known experimental conditions which permit optimal enzymatic activity. For example, assays for N-acetylglutamate kinase are presented by Wolf, E. C. and Weiss, R. L. (1980) *J. Biol. Chem. 255*:9189-9195.

CLAIMS

What is claimed is:

1. An isolated nucleic acid fragment encoding an N-acetylglutamate kinase comprising a member selected from the group consisting of:

(a) an isolated nucleic acid fragment encoding an amino acid sequence that is at least 95% identical to the amino acid sequence set forth in a member selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6 and SEQ ID NO:8 and encoding a functional enzyme;

(b) an isolated nucleic acid fragment that is complementary to (a).

2. The isolated nucleic acid fragment of Claim 1 wherein nucleic acid fragment is a functional RNA.

3. The isolated nucleic acid fragment of Claim 1 wherein the nucleotide sequence of the fragment comprises the sequence set forth in a member selected from the group consisting of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5 and SEQ ID NO:7.

4. A chimeric gene comprising the nucleic acid fragment of Claim 1 operably linked to suitable regulatory sequences.

5. A transformed host cell comprising the chimeric gene of Claim 4.

6. An isolated nucleic acid fragment encoding an N-acetylglutamate kinase having the sequence set forth in SEQ ID NO:12, where:

Xaa at position 2 is Leu or Met,
Xaa at position 3 is Leu or Ala,
Xaa at position 4 or 176 is Thr, Ala or Gly,
Xaa at position 5 is Pro or none,
Xaa at position 6 is Trp, His or none,
Xaa at position 7 is Leu or none,
Xaa at position 8, 9, or 46 is Ser, Ala or none,
Xaa at position 10 is Ser, Lys or Pro,
Xaa at position 11 is Lys, Ser, Thr or Ala,
Xaa at position 12 Leu or Phe,
Xaa at position 13 is Pro, Thr or none,
Xaa at position 14 is Val, Leu or Asn,
Xaa at position 15 is Pro or Leu,
Xaa at position 16 or 182 is Ser or Cys,
Xaa at position 17 or 36 is Pro, Thr or Ala,
Xaa at position 18 is Pro, Arg or Ser,
Xaa at position 19 is Pro, Val, Phe or Leu,
Xaa at position 20 or 30 is Ser, Lys, Thr or Arg,

Xaa at position 21 is Gly, Phe or Asn,
Xaa at position 23 or 50 is Thr, Ala or Asn,
Xaa at position 24 is Leu, Pro or Lys,
Xaa at position 25 is Ser, Gly, Pro or Lys,
Xaa at position 26 is Ser, Pro, Gln or Ala,
Xaa at position 27 is Asn or Ala,
Xaa at position 28 is His, Gln or Arg,
Xaa at position 29 is Ala, Leu or Val,
Xaa at position 31 is Pro or Thr,
Xaa at position 32 is Leu, Ile or Ser,
Xaa at position 33 is Ala or His,
Xaa at position 34, 49, 51, 52, 74, 194, 252, 256 or 309 is Ala or Ser,
Xaa at position 35 is Pro, Ser or Phe,
Xaa at position 37 is Cys, Ala, Ser or Pro,
Xaa at position 38 is Arg, Pro, Thr or His,
Xaa at position 39 is Arg, Gly or none,
Xaa at position 40 or 43 is Arg or Leu,
Xaa at position 41 is Ser or Arg,
Xaa at position 42 is Arg, Cys, His or Gly,
Xaa at position 44 is Arg or Ala,
Xaa at position 45 or 253 is Ile, Leu or Val,
Xaa at position 47 is Ala, Val or none,
Xaa at position 48 is Thr, Val or none,
Xaa at position 53 is Pro, Gln or Ser,
Xaa at position 54 or 57 is Ser, Ala or Pro,
Xaa at position 55 is Pro, Ala or Leu,
Xaa at position 56 is Ser, Leu or Ala,
Xaa at position 59 is Ala, Glu or Gln,
Xaa at position 60, 227 or 428 is Ala or Thr,
Xaa at position 62 is Thr, Ser or none,
Xaa at position 63, 191, 197, 269 or 291 is Glu or Ala,
Xaa at position 64 is Ala, Ser or Gly,
Xaa at position 65 id Leu or Gln,
Xaa at position 66 is Ser, Tyr or Asn,
Xaa at position 80 or 82 is Arg or Lys,
Xaa at position 86, 137, 154, 179, 226, 281 is Val or Ile,
Xaa at position 106 is Arg or Asn,
Xaa at position 117, 215 or 308 is Arg or His.

Xaa at position 133 is Leu, Gly or Gln,
Xaa at position 135 or 209 is Val or Lys,
Xaa at position 136 is Gly or Asn,
Xaa at position 138 is Glu or Pro,
Xaa at position 139 is Gln or Val,
Xaa at position 142, 210 or 272 is Asp or Asn,
Xaa at position 150 is Leu or Asp,
Xaa at position 151 or 190 is Thr or Asn,
Xaa at position 156 is Glu or Ser,
Xaa at position 161 or 201 is Gly or Arg,
Xaa at position 166 is Asn, Glu, Thr or Gln,
Xaa at position 172 is Asn, Lys or Arg,
Xaa at position 173 is Ile, Leu, Lys or Pro,
Xaa at position 174, 211 or 237 is Ala or Pro,
Xaa at position 177, 228 or 234 is Thr or Ser,
Xaa at position 180 or 221 is Gly or Ser,
Xaa at position 183 is Gly, Trp or Arg,
Xaa at position 184 is Lys or Met,
Xaa at position 185 or 292 is Asp or Glu,
Xaa at position 186, 230 or 296 is Gly or Ala,
Xaa at position 189, 217, 249, 259 or 261 is Ile or Leu,
Xaa at position 196 is Asn, Lys or Asp,
Xaa at position 198 or 284 is Lys or Ala,
Xaa at position 199 is Ala, Gly or Asp,
Xaa at position 202 is Phe or Tyr,
Xaa at position 205 is Glu or Gly,
Xaa at position 207 is Ser, Trp, Ala or Thr,
Xaa at position 212 or 220 is Thr, Ala or Ser,
Xaa at position 216 is Pro or Ser,
Xaa at position 219 is Ala or Asp,
Xaa at position 245 is Ala or Val,
Xaa at position 275 is Gly or Asp,
Xaa at position 280 is Val, Glu or Lys,
Xaa at position 288 is Lys, Gln or Arg,
Xaa at position 290 is Val or Met,
Xaa at position 294 is Lys or Gln,
Xaa at position 304 is Glu, Asn or Gly,
and Xaa at position 405 is His or Ile.

7. A method for evaluating at least one compound for its ability to inhibit the activity of an N-acetylglutamate kinase, the method comprising the steps of:

(a) transforming a host cell with a chimeric gene comprising a nucleic acid fragment encoding an N-acetylglutamate kinase, operably linked to suitable regulatory sequences;

(b) growing the transformed host cell under conditions that are suitable for expression of the chimeric gene wherein expression of the chimeric gene results in production of the N-acetylglutamate kinase encoded by the operably linked nucleic acid fragment in the transformed host cell;

(c) optionally purifying the N-acetylglutamate kinase expressed by the transformed host cell;

(d) treating the N-acetylglutamate kinase with a compound to be tested; and

(e) comparing the activity of the N-acetylglutamate kinase that has been treated with a test compound to the activity of an untreated N-acetylglutamate kinase,

thereby selecting compounds with potential for inhibitory activity.

FIGURE 1

```
              *                                                         +
SEQ ID NO:11  MSSTQDYIGE--------------------------------------------------
SEQ ID NO:02  MLLTKPYLSNSLLPVPSPPPSGPTLSSNHASPLAAPTCR-RSRLRISATSTAAPSPSSAA
SEQ ID NO:04  MLLAKPHLSSSSF-LPSTRVSSPAPGPNHAKPIAASPAP-RRCLRLAVTSAAAPAASSAE
SEQ ID NO:06  MMAG----AAKTLTNLCPSFPFPTKPQNQLTTSHAFPSTRLRHRAISAVANAAQPPLAAA
SEQ ID NO:08  MLLTKPH---PALTLPSASLPNPNLKAARVRPLASSAPHGRRGLRV---SASSSSLAPAQ
SEQ ID NO:12  MXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXAX
              1                                                          60

               +     **++*** **+** * *+* **********+++*+++** *+*+++++*+ **
SEQ ID NO:11  ---EAATRVKILSEALPYIQHFAGRTVVVKYGGAAMKDSNLKDKVIRDIVFMASVGIRPV
SEQ ID NO:02  AATASLSRVDVLSEALPFIQRFKGKTVVVKYGGAAMKSPELQASVIRDLVLLSCVGLRPV
SEQ ID NO:04  AA-AALSRVDVLSEALPFIQRFKGKTVVVKYGGAAMKSPELQASVIRDLVLLSCVGLHPV
SEQ ID NO:06  TATEGQYRVDVLSESLPFIQKFRGKTIVVKYGGAAMKSPELQASVINDLVLLSCVGLRPV
SEQ ID NO:08  AASAALNRVDVLSEALPFIQRFKGKTVVVKYGGAAMKSPELQASVIRDLVLLSCVGLRPV
SEQ ID NO:12   AXXXXXRVDVLSEXLPFIQXFXGKTXVVKYGGAAMKSPELQASVIXDLVLLSCVGLXPV
              61                                                        120

              +**********+** +      *+ *****                 * +**   *  *
SEQ ID NO:11  VVHGGGPEINTWLDKVGIEPQFKDGLRVTDAATMDIVEMVLVGRVNKELVNLINQAGGKA
SEQ ID NO:02  LVHGGGPEINSWLLRVGVEPQFRDGLRVTDALTMEVVEMVLVGKVNKNLVSLINIAGGTA
SEQ ID NO:04  LVHGGGPEINSWLLRVGVEPQFRNGLRVTDALNMEVVEMVLVRKVNKELLSLIKLPGGSA
SEQ ID NO:06  LVHGGGPEINSWLGRLNIPAVFRDGLRVTDADTMEIVSMVLVGKVNKTLVSLINKAGATA
SEQ ID NO:08  LVHGGGPEINSWLQRVGVXPQFRNGLRVTXXXXXXXXXXXXXXXXXKQLLSLIRPAGTTA
SEQ ID NO:12  LVHGGGPEINSWLXRXXXXXXFRXGLRVTDAXXMEXVXMVLVXKVNKXLXSLIXXXGXXA
              121                                                       180
```

FIGURE 1 (CONTINUED)

```
                  *     +    *+ +    +  ** * +    **+   +   +***   * *** ** +
SEQ ID NO:11  VGLCGKDGQLMTARTMTN-KDVGFVGEVSSVDARVVETLVKSGYIPVISSVAADEFGQAH
SEQ ID NO:02  IGLCGKDARLITARPSPNAAALGFVGEVSRVDATVLHPIIAAGHIPVIATVAADETGQAY
SEQ ID NO:04  VSLCWKEARLLNERPSPXEKGLRFVGGVWRVDATVLHPIIASGHIPVIATVGADETGQAY
SEQ ID NO:06  VGLSGMDGRLLTARPAPKAADLGYVGEVARVDPAVLRSLIDTSHIPVVTSVAADESGQPY
SEQ ID NO:08  VGLCRKDGRILTERPSPDAAALGFVGEVTRKNPSVLHPIIASSHIPVIATVAADETGQAY
SEQ ID NO:12  XXLXXXXXRLXXXRPXPXXXXLXXVGXVXRXXXXVLXXXIXXXHIPVXXXVXADEXGQXY
              181                                                      240

              ****** *** **  ***** * **+ *** *+ ** +*++  ** ++  +   * * **
SEQ ID NO:11  NINADTCAGELAAALGAEKLILLLTDTRGILRDYKDPSTLIHKLDIQQARELIGSGIVAGG
SEQ ID NO:02  NINADTAAGEIAAAVGAEKLLLLTDVSGILADRNDPGSLVKVVDIAGVRKMVADGKVAGG
SEQ ID NO:04  NINADTAAGEIAAAVGAEKLLLLTDVSGILADRNDPGSLVKEIDIAGVRQMVADGQVAGG
SEQ ID NO:06  NINADTVAGELAASLGAEKLILLTDVAGILEDRNDPDSLVKKIDIKGVKKMMEDGKVGGG
SEQ ID NO:08  NINADTAAGEIAAAIGAEKLLLITDVSGILADRDDPGSLVKEIDIAGVRRMVAEGKVGGG
SEQ ID NO:12  NINADTXAGEXAAXXGAEKLXLXTDVXGILXDRXDPXSLVKXXDIXGVRXMXXXGXVXGG
              241                                                      300

              ***** ***  ***** +*+*+***+**+****+*+**+* *+***++
SEQ ID NO:11  MIPKVTCCVRSLAQGVRAAHILDGRLPHALLLEVFTDLGIGSMIVASGYDL
SEQ ID NO:02  MIPKVECCVHALAQGVHTASIIDGRVPHSLLLEILTDEGTGTMITG.----
SEQ ID NO:04  MIPKVECCVRALAQGVHTASIIDGRVPHSLLLEILTDEGTGTMITG.----
SEQ ID NO:06  MIPKVNCCVRSLAQGVITASIIDGRVPHSLLLEILTDEGAGTMITG.----
SEQ ID NO:08  MIPKVGCCVRALAQGVHTASIIDGRVPHSLLLEILTDEGTGTMITG.----
SEQ ID NO:12  MIPKVXCCVXXLAQGVXTASILDGRVPHSLLLEILTDEGXGTMITG
              301                                          351
```

International Application No
PCT/US 99/15931

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N15/54 C12N15/82 C12N9/12 C12N5/10 C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12N C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	REITH M ET AL: "Two amino-acid biosynthetic genes are encoded on the plastid genome of the red alga Porphyra umbilicalis." CURRENT GENETICS, (1993 JAN) 23 (1) 59-65. , XP002118025 cited in the application the whole document ---	1-6
A	MCKAY G AND SHARGOOL P.: "Purification and characterization of N-acetylglutamate 5-phosphotransferase from pea (Pisum sativum) cotyledons" BIOCHEMICAL JOURNAL, vol. 195, no. 1, 1981, pages 71-81, XP002118026 see the whole document; for claim 7 see esp. p.73 r. col. --- -/--	1-7

[X] Further documents are listed in the continuation of box C. [X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
7 October 1999	21/10/1999

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Kania, T

Form PCT/ISA/210 (second sheet) (July 1992)

Information on patent family members

International Application No
PCT/US 99/15931

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 0261627 A	30-03-1988	JP 1994602 C	22-11-1995
		JP 7028749 B	05-04-1995
		JP 63079597 A	09-04-1988
		DE 3785530 A	27-05-1993
		DE 3785530 T	18-11-1993
		US 5017482 A	21-05-1991

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

SEQUENCE LISTING

<110> Cahoon, Rebecca E.
Gutteridge, Steven
Lee, Jian Ming
Rafalski, Antoni

<120> Ornithine Biosynthesis Enzymes

<130> BB-1174

<140>
<141>

<150> 60/093,209
<151> July 17, 1998

<160> 12

<170> Microsoft Office 97

<210> 1
<211> 1201
<212> DNA
<213> Zea mays

<400> 1
tcgagctcga gctcgagccc cagtcaccgc agccatgctc ctcacgaaac cctacctctc 60
caactcgctc cttccagtcc catccccgcc gccgtcgggc cctactctca gctccaacca 120
tgcaagcccc cttgccgccc ctacttgccg tcgcagccgc ctccgcatct ccgccacatc 180
cacggctgcg ccgtctcctt cgtcggctgc cgctgccacc gcgtcgctga gtcgagtgga 240
cgtgctctcg gaggcgctcc cttttattca gcgattcaaa ggcaagacgg tggtggtcaa 300
gtacggcggt gcggcgatga agtccccgga gctgcaggcg tccgtgatcc gcgatctcgt 360
gctgctctcc tgcgtcggcc tccgccccgt gcttgttcac ggcggcggtc cggagattaa 420
ttcctggctg ctgcgcgtcg gcgtcgagcc gcagttccgc gacggcctcc gcgtcacgga 480
cgcgctcacc atggaggtcg tcgagatggt gctagtcggg aaggtcaaca aaaaccttgt 540
ttccctcatc aacatcgcgg gaggcaccgc cattggtctg tgcggcaagg acgcgcgcct 600
tatcaccgct cgcccgtctc caaatgcagc ggcgctggga ttcgtcggcg aggtttcgcg 660
cgtggacgcc accgtcctcc atcccatcat cgccgcgggc catatcccgg ttatcgccac 720
cgttgccgcc gacgagactg ggcaagccta taacatcaat gctgatacgg cggctggcga 780
gattgccgct gccgtgggcg ccgagaagct gctgttgctc acagatgtgt ctggcatttt 840
ggcggaccgt aatgaccctg ggagcctggt gaaggtggtc gacattgctg cggtgcggaa 900
gatggtggct gacggcaagg tagctggtgg gatgataccc aaggtggagt gttgtgttca 960
cgcccttgca caaggtgtac acaccgcaag tatcattgat gggcgtgttc cacactctct 1020
tctgcttgag attctcacag acgagggcac aggcaccatg atcactggct cagctgcttc 1080
atgccttcat ggtattttcc tgtgcctctt ttctcatatt gttgtgtttt atggctatgt 1140
agactaaact caagattgca ataagactac ctaagtttgg ttgaaaaaaa aaaaaaaaaa 1200
a 1201

<210> 2
<211> 345
<212> PRT
<213> Zea mays

<400> 2
Met Leu Leu Thr Lys Pro Tyr Leu Ser Asn Ser Leu Leu Pro Val Pro
1 5 10 15

Ser Pro Pro Pro Ser Gly Pro Thr Leu Ser Ser Asn His Ala Ser Pro
20 25 30

```
Leu Ala Ala Pro Thr Cys Arg Arg Ser Arg Leu Arg Ile Ser Ala Thr
        35                  40                  45

Ser Thr Ala Ala Pro Ser Pro Ser Ser Ala Ala Ala Ala Thr Ala Ser
    50                  55                  60

Leu Ser Arg Val Asp Val Leu Ser Glu Ala Leu Pro Phe Ile Gln Arg
65                  70                  75                  80

Phe Lys Gly Lys Thr Val Val Val Lys Tyr Gly Gly Ala Ala Met Lys
                85                  90                  95

Ser Pro Glu Leu Gln Ala Ser Val Ile Arg Asp Leu Val Leu Leu Ser
            100                 105                 110

Cys Val Gly Leu Arg Pro Val Leu Val His Gly Gly Gly Pro Glu Ile
        115                 120                 125

Asn Ser Trp Leu Leu Arg Val Gly Val Glu Pro Gln Phe Arg Asp Gly
    130                 135                 140

Leu Arg Val Thr Asp Ala Leu Thr Met Glu Val Val Glu Met Val Leu
145                 150                 155                 160

Val Gly Lys Val Asn Lys Asn Leu Val Ser Leu Ile Asn Ile Ala Gly
                165                 170                 175

Gly Thr Ala Ile Gly Leu Cys Gly Lys Asp Ala Arg Leu Ile Thr Ala
            180                 185                 190

Arg Pro Ser Pro Asn Ala Ala Ala Leu Gly Phe Val Gly Glu Val Ser
        195                 200                 205

Arg Val Asp Ala Thr Val Leu His Pro Ile Ile Ala Ala Gly His Ile
    210                 215                 220

Pro Val Ile Ala Thr Val Ala Ala Asp Glu Thr Gly Gln Ala Tyr Asn
225                 230                 235                 240

Ile Asn Ala Asp Thr Ala Ala Gly Glu Ile Ala Ala Ala Val Gly Ala
                245                 250                 255

Glu Lys Leu Leu Leu Leu Thr Asp Val Ser Gly Ile Leu Ala Asp Arg
            260                 265                 270

Asn Asp Pro Gly Ser Leu Val Lys Val Val Asp Ile Ala Gly Val Arg
        275                 280                 285

Lys Met Val Ala Asp Gly Lys Val Ala Gly Gly Met Ile Pro Lys Val
    290                 295                 300

Glu Cys Cys Val His Ala Leu Ala Gln Gly Val His Thr Ala Ser Ile
305                 310                 315                 320

Ile Asp Gly Arg Val Pro His Ser Leu Leu Leu Glu Ile Leu Thr Asp
                325                 330                 335

Glu Gly Thr Gly Thr Met Ile Thr Gly
            340                 345
```

<210> 3
<211> 1186
<212> DNA
<213> Oryza sativa

<400> 3
gcacgagtac agcgccgccg ccgccgccat gctcctcgcg aagccccacc tctcctcctc 60
ctctttcctc ccatccacgc gggtgtctag ccccgctccg ggtcccaacc acgcaaagcc 120
catcgccgcc tctcccgccc ctcgacgctg cctccgtctc gccgtcacat ccgccgcggc 180
gccggctgct tcgtcggcgg aggcggcggc ggcgctgagc cgcgtggatg tgctctcaga 240
ggcgctcccc ttcatccagc gcttcaaggg gaagaccgtg gtggtgaagt acggcggcgc 300
ggcgatgaag tcgcccggagc tccaggcttc agtgatccgc gacctggtcc tcctctcgtg 360
cgtcggcctc cacccccgtgc tcgtccacgg cggcgggccc gagatcaact cctggctgct 420
ccgcgtcggc gtcgagccgc agttccggaa cggcctccgc gtcactgacg cgctcaacat 480
ggaggtcgtc gagatggtgc tcgtccgcaa ggtcaacaaa gaactcctct ccctcatcaa 540
actcccgggg gggagccgccg taagtctctg ttggaaggaa gctcgcctcc tcaacgagcg 600
gccctccccg aangaaaagg gccttcggtt tgtcggcggg gtctggcgcg tggacgccac 660
cgtcctccac ccaatcatcg cctccggtca catcccggtc atcgccactg tgggcgccga 720
cgagaccggg caggcctaca acatcaacgc tgacacggcg gccggcgaga tcgccgccgc 780
ggtcggcgcg gagaagctgt tgctgctcac agatgtgtct ggaattctgg acgaccgtaa 840
tgaccccggg agtctggtga aagagatcga cattgctggg gtgcggcaga tggtggccga 900
cgggcaggta gctggtggga tgataccgaa ggtggaatgc tgcgtgcgtg ccctcgcaca 960
gggcgtgcac actgcaagca tcatcgatgg gcgtgtcccg cactcgttgc tgctcgagat 1020
tctcacagat gagggcactg gcactatgat cactggctga ggtgattcat cccgtcgtgg 1080
tattctccgg tgcctttctt ctcatactgt aatgtaattt gcatttgata tgcctcatga 1140
ttgcaataag aattgtattc ctcaaaaaaa aaaaaaaaaa aaaaaa 1186

<210> 4
<211> 343
<212> PRT
<213> Oryza sativa

<220>
<221> UNSURE
<222> (195)

<400> 4
Met Leu Leu Ala Lys Pro His Leu Ser Ser Ser Ser Phe Leu Pro Ser
1 5 10 15

Thr Arg Val Ser Ser Pro Ala Pro Gly Pro Asn His Ala Lys Pro Ile
20 25 30

Ala Ala Ser Pro Ala Pro Arg Arg Cys Leu Arg Leu Ala Val Thr Ser
35 40 45

Ala Ala Ala Pro Ala Ala Ser Ser Ala Glu Ala Ala Ala Ala Leu Ser
50 55 60

Arg Val Asp Val Leu Ser Glu Ala Leu Pro Phe Ile Gln Arg Phe Lys
65 70 75 80

Gly Lys Thr Val Val Val Lys Tyr Gly Gly Ala Ala Met Lys Ser Pro
85 90 95

Glu Leu Gln Ala Ser Val Ile Arg Asp Leu Val Leu Leu Ser Cys Val
100 105 110

```
Gly Leu His Pro Val Leu Val His Gly Gly Gly Pro Glu Ile Asn Ser
        115                 120                 125

Trp Leu Leu Arg Val Gly Val Glu Pro Gln Phe Arg Asn Gly Leu Arg
    130                 135                 140

Val Thr Asp Ala Leu Asn Met Glu Val Val Glu Met Val Leu Val Arg
145                 150                 155                 160

Lys Val Asn Lys Glu Leu Leu Ser Leu Ile Lys Leu Pro Gly Gly Ser
                165                 170                 175

Ala Val Ser Leu Cys Trp Lys Glu Ala Arg Leu Leu Asn Glu Arg Pro
            180                 185                 190

Ser Pro Xaa Glu Lys Gly Leu Arg Phe Val Gly Gly Val Trp Arg Val
        195                 200                 205

Asp Ala Thr Val Leu His Pro Ile Ile Ala Ser Gly His Ile Pro Val
    210                 215                 220

Ile Ala Thr Val Gly Ala Asp Glu Thr Gly Gln Ala Tyr Asn Ile Asn
225                 230                 235                 240

Ala Asp Thr Ala Ala Gly Glu Ile Ala Ala Ala Val Gly Ala Glu Lys
                245                 250                 255

Leu Leu Leu Leu Thr Asp Val Ser Gly Ile Leu Ala Asp Arg Asn Asp
            260                 265                 270

Pro Gly Ser Leu Val Lys Glu Ile Asp Ile Ala Gly Val Arg Gln Met
        275                 280                 285

Val Ala Asp Gly Gln Val Ala Gly Gly Met Ile Pro Lys Val Glu Cys
    290                 295                 300

Cys Val Arg Ala Leu Ala Gln Gly Val His Thr Ala Ser Ile Ile Asp
305                 310                 315                 320

Gly Arg Val Pro His Ser Leu Leu Leu Glu Ile Leu Thr Asp Glu Gly
                325                 330                 335

Thr Gly Thr Met Ile Thr Gly
            340
```

<210> 5
<211> 1204
<212> DNA
<213> Glycine max

<400> 5

```
gcacgagatg atggcaggtg cagccaaaac cctaaccaat ctttgccccc ctttcccatt   60
cccaaccaaa ccccaaaacc aactcaccac tagccacgct ttcccttcca ctcgcctccg  120
ccaccgcgcc atttccgcgg tggcgaacgc ggcgcaacct ccactcgccg ccgccactgc  180
caccgagggt cagtaccgag tcgatgtgct ctcggagtcg ctcccccttca tccagaaatt  240
ccgcggcaaa accatcgtcg tcaagtacgg cggcgccgcc atgaagtccc cggagctcca  300
ggcctccgtg atcaacgacc ttgtcctcct ctcctgcgtc ggcctccgcc ccgtcctggt  360
ccacggcggc ggccccgaga tcaactcctg gctcggccgc ctcaacatcc ccgccgtctt  420
ccgcgacggc ctccccgtca ccgacgccga caccatggag atcgtctcca tggtcctcgt  480
cggaaaagtc aacaaaaccc tagtttctct aattaacaag gccggcgcca ccgccgtcgg  540
```

cctctctggc atggacggcc gcctcctcac cgcccgcccc gctcccaagc ccgccgacct 600
cggctacgtc ggcgaggtcg cacgcgtcga tcccgccgtc ctccgctccc taatcgacac 660
cagccacatc cccgtcgtca cctccgtcgc cgccgatgaa tccggacagc cctacaacat 720
caacgccgac accgtcgccg gagaattggc agcgtcgctc ggcgcggaga agctgattct 780
gctgaccgat gtggcgggaa ttctggaaga tcggaacgac cctgacagct tggtgaagaa 840
gattgacata aaaggagtga agaaaatgat ggaagatgga aaagttggtg gtggaatgat 900
acctaaggtt aattgttgcg ttaggtcctt ggcgcaaggg gttattacag cgagtattat 960
tgatggtagg gttccgcatt ctttgttgct tgagattttg actgatgaag gtgctggaac 1020
tatgataact ggataagttt atttatttat ggtgtttgga ttttttcttt tcaatcaagc 1080
cttgagttga ggttgcattg cagcacttgt tttgttagag attggtgatt gtttttaagt 1140
gcgtgtaatg tgagagatgg ttgaattgaa ttgaatgttt cagaaaaaaa aaaaaaaaaa 1200
aaaa 1204

<210> 6
<211> 342
<212> PRT
<213> Glycine max

<400> 6

Met Met Ala Gly Ala Ala Lys Thr Leu Thr Asn Leu Cys Pro Ser Phe
1 5 10 15

Pro Phe Pro Thr Lys Pro Gln Asn Gln Leu Thr Thr Ser His Ala Phe
20 25 30

Pro Ser Thr Arg Leu Arg His Arg Ala Ile Ser Ala Val Ala Asn Ala
35 40 45

Ala Gln Pro Pro Leu Ala Ala Ala Thr Ala Thr Glu Gly Gln Tyr Arg
50 55 60

Val Asp Val Leu Ser Glu Ser Leu Pro Phe Ile Gln Lys Phe Arg Gly
65 70 75 80

Lys Thr Ile Val Val Lys Tyr Gly Gly Ala Ala Met Lys Ser Pro Glu
85 90 95

Leu Gln Ala Ser Val Ile Asn Asp Leu Val Leu Leu Ser Cys Val Gly
100 105 110

Leu Arg Pro Val Leu Val His Gly Gly Gly Pro Glu Ile Asn Ser Trp
115 120 125

Leu Gly Arg Leu Asn Ile Pro Ala Val Phe Arg Asp Gly Leu Arg Val
130 135 140

Thr Asp Ala Asp Thr Met Glu Ile Val Ser Met Val Leu Val Gly Lys
145 150 155 160

Val Asn Lys Thr Leu Val Ser Leu Ile Asn Lys Ala Gly Ala Thr Ala
165 170 175

Val Gly Leu Ser Gly Met Asp Gly Arg Leu Leu Thr Ala Arg Pro Ala
180 185 190

Pro Lys Ala Ala Asp Leu Gly Tyr Val Gly Glu Val Ala Arg Val Asp
195 200 205

Pro Ala Val Leu Arg Ser Leu Ile Asp Thr Ser His Ile Pro Val Val
210 215 220

```
Thr Ser Val Ala Ala Asp Glu Ser Gly Gln Pro Tyr Asn Ile Asn Ala
225                 230                 235                 240

Asp Thr Val Ala Gly Glu Leu Ala Ala Ser Leu Gly Ala Glu Lys Leu
                245                 250                 255

Ile Leu Leu Thr Asp Val Ala Gly Ile Leu Glu Asp Arg Asn Asp Pro
            260                 265                 270

Asp Ser Leu Val Lys Lys Ile Asp Ile Lys Gly Val Lys Lys Met Met
        275                 280                 285

Glu Asp Gly Lys Val Gly Gly Gly Met Ile Pro Lys Val Asn Cys Cys
    290                 295                 300

Val Arg Ser Leu Ala Gln Gly Val Ile Thr Ala Ser Ile Ile Asp Gly
305                 310                 315                 320

Arg Val Pro His Ser Leu Leu Leu Glu Ile Leu Thr Asp Glu Gly Ala
                325                 330                 335

Gly Thr Met Ile Thr Gly
            340
```

<210> 7
<211> 1246
<212> DNA
<213> Triticum aestivum

<220>
<221> unsure
<222> (492)..(542)

<400> 7

```
gcacgaggtc taggcaagca acccggcggc gcgaacacca agtccccccg gcggccgcag   60
caatgctcct aaccaagccc caccccgccc tcaccctccc ctccgcatcc ctcccaaatc  120
ctaacctaaa ggccgcccgc gtcaggcccc tcgcctcctc cgcgccccat ggacgccgcg  180
ggctccgcgt ctcggcctcc tcctcctccc tggcgccagc gcaggccgcg tccgcggcgc  240
tgaaccgcgt ggacgtcctg tcggaggcgc tccccttcat ccagcggttc aagggcaaga  300
cggtggtggt caagtacggc ggcgcggcca tgaagtcgcc ggagctgcag gcgtcggtga  360
tccgcgacct ggtcctcctc tcctgcgtcg gcctgcgccc cgtgctcgtg cacggcggcg  420
gcccggagat caactcctgg ctgcagcgcg tcggggtcta gccgcagttc cgcaacggcc  480
tccgcgtcac gnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn  540
nnaagcagct cttatcccta atcaggcctg cggggaccac agcagttggc ctctgcagaa  600
aggacgggcg catcctaacg gagcgccccc ccccagacgc cgcagccctc gggttcgtcg  660
gcgaggtcac gagaaaaaac ccctctgtgc tccacccgat catcgcctcc agccacatcc  720
cggtcatcgc caccctggct gccgacgaga ccgggcaagc ctataacatc aacgctgaca  780
ctgcggcggg ggagatcgcg gctgccattg gcgccgagaa gctgttgctg atcactgacg  840
tgtccggcat actcgcggac cgggatgacc ccgggagcct ggtgaaggag attgacatcg  900
ccggcgtacg gcggatggtg gccgagggca aggtgggtgg gggcatgata cccaaggtgg  960
ggtgctgcgt gcgcgcgctg gcgcagggcg tgcacacggc cagcatcatt gacggccgcg 1020
tcccgcactc tctcctgctc gaaatcctca ccgacgaggg caccggcacc atgatcaccg 1080
gctgaaactt gtttgtttgt tgttgttttt ttcttttctt ttttggttca cattctttgg 1140
gttggatggt tttcgatccc tcatttgtgt taaatgtgtt ttcgattcga atcctgaaca 1200
aggagtggtg aaacattgca gctttaagca aaaaaaaaaa aaaaaa                 1246
```

<210> 8
<211> 340

<212> PRT
<213> Triticum aestivum

<220>
<221> UNSURE
<222> (133)

<220>
<221> UNSURE
<222> (144)..(160)

<400> 8

```
Met Leu Leu Thr Lys Pro His Pro Ala Leu Thr Leu Pro Ser Ala Ser
 1               5                  10                  15

Leu Pro Asn Pro Asn Leu Lys Ala Ala Arg Val Arg Pro Leu Ala Ser
            20                  25                  30

Ser Ala Pro His Gly Arg Arg Gly Leu Arg Val Ser Ala Ser Ser Ser
        35                  40                  45

Ser Leu Ala Pro Ala Gln Ala Ala Ser Ala Ala Leu Asn Arg Val Asp
    50                  55                  60

Val Leu Ser Glu Ala Leu Pro Phe Ile Gln Arg Phe Lys Gly Lys Thr
65                  70                  75                  80

Val Val Val Lys Tyr Gly Gly Ala Ala Met Lys Ser Pro Glu Leu Gln
                85                  90                  95

Ala Ser Val Ile Arg Asp Leu Val Leu Leu Ser Cys Val Gly Leu Arg
            100                 105                 110

Pro Val Leu Val His Gly Gly Gly Pro Glu Ile Asn Ser Trp Leu Gln
        115                 120                 125

Arg Val Gly Val Xaa Pro Gln Phe Arg Asn Gly Leu Arg Val Thr Xaa
    130                 135                 140

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
145                 150                 155                 160

Lys Gln Leu Leu Ser Leu Ile Arg Pro Ala Gly Thr Thr Ala Val Gly
                165                 170                 175

Leu Cys Arg Lys Asp Gly Arg Ile Leu Thr Glu Arg Pro Ser Pro Asp
            180                 185                 190

Ala Ala Ala Leu Gly Phe Val Gly Glu Val Thr Arg Lys Asn Pro Ser
        195                 200                 205

Val Leu His Pro Ile Ile Ala Ser Ser His Ile Pro Val Ile Ala Thr
    210                 215                 220

Val Ala Ala Asp Glu Thr Gly Gln Ala Tyr Asn Ile Asn Ala Asp Thr
225                 230                 235                 240

Ala Ala Gly Glu Ile Ala Ala Ala Ile Gly Ala Glu Lys Leu Leu Leu
                245                 250                 255
```

```
Ile Thr Asp Val Ser Gly Ile Leu Ala Asp Arg Asp Asp Pro Gly Ser
            260                 265                 270

Leu Val Lys Glu Ile Asp Ile Ala Gly Val Arg Arg Met Val Ala Glu
        275                 280                 285

Gly Lys Val Gly Gly Gly Met Ile Pro Lys Val Gly Cys Cys Val Arg
    290                 295                 300

Ala Leu Ala Gln Gly Val His Thr Ala Ser Ile Ile Asp Gly Arg Val
305                 310                 315                 320

Pro His Ser Leu Leu Leu Glu Ile Leu Thr Asp Glu Gly Thr Gly Thr
                325                 330                 335

Met Ile Thr Gly
            340
```

<210> 9
<211> 439
<212> DNA
<213> Triticum aestivum

<400> 9

```
gcacgaggtg agattgccgc tgcggtgggc gccgagaaat tgctactgct cacagatgtg  60
tctgggatac tggcggaccg taatgaccct ggcagcctgg tgaaggagat tgacatcgct 120
ggggtgcggc agatggtatc cggtgggcag gttgctggtg gaatgatccc aaaggtggag 180
tgctgcgtga gagccctcgc ccagggtgtg cacactgcaa gcatcatcga tgggcgtgtc 240
ccgcactcgc tgttgctcga gattctcaca gatgagggca ctggcacaat gatcaccggc 300
taaggtgtaa aatgcctcct tggtacttcc ttatgccttt ctgttcatac tgccaatctg 360
ccatgtaatt tatgccaatg tagcctcacc tcatgattgc aataagagta ccttcctgac 420
aaaaaaaaaa aaaaaaaaa                                              439
```

<210> 10
<211> 100
<212> PRT
<213> Triticum aestivum

<400> 10

```
Ala Arg Gly Glu Ile Ala Ala Ala Val Gly Ala Glu Lys Leu Leu Leu
  1               5                  10                  15

Leu Thr Asp Val Ser Gly Ile Leu Ala Asp Arg Asn Asp Pro Gly Ser
             20                  25                  30

Leu Val Lys Glu Ile Asp Ile Ala Gly Val Arg Gln Met Val Ser Gly
         35                  40                  45

Gly Gln Val Ala Gly Gly Met Ile Pro Lys Val Glu Cys Cys Val Arg
     50                  55                  60

Ala Leu Ala Gln Gly Val His Thr Ala Ser Ile Ile Asp Gly Arg Val
 65                  70                  75                  80

Pro His Ser Leu Leu Leu Glu Ile Leu Thr Asp Glu Gly Thr Gly Thr
                 85                  90                  95

Met Ile Thr Gly
            100
```

<210> 11
<211> 297
<212> PRT
<213> Synechocystis sp.

<400> 11

```
Met Ser Ser Thr Gln Asp Tyr Ile Gly Glu Glu Ala Ala Thr Arg Val
 1               5                  10                  15

Lys Ile Leu Ser Glu Ala Leu Pro Tyr Ile Gln His Phe Ala Gly Arg
            20                  25                  30

Thr Val Val Val Lys Tyr Gly Gly Ala Ala Met Lys Asp Ser Asn Leu
        35                  40                  45

Lys Asp Lys Val Ile Arg Asp Ile Val Phe Met Ala Ser Val Gly Ile
    50                  55                  60

Arg Pro Val Val Val His Gly Gly Gly Pro Glu Ile Asn Thr Trp Leu
65                  70                  75                  80

Asp Lys Val Gly Ile Glu Pro Gln Phe Lys Asp Gly Leu Arg Val Thr
                85                  90                  95

Asp Ala Ala Thr Met Asp Ile Val Glu Met Val Leu Val Gly Arg Val
            100                 105                 110

Asn Lys Glu Leu Val Asn Leu Ile Asn Gln Ala Gly Gly Lys Ala Val
        115                 120                 125

Gly Leu Cys Gly Lys Asp Gly Gln Leu Met Thr Ala Arg Thr Met Thr
    130                 135                 140

Asn Lys Asp Val Gly Phe Val Gly Glu Val Ser Ser Val Asp Ala Arg
145                 150                 155                 160

Val Val Glu Thr Leu Val Lys Ser Gly Tyr Ile Pro Val Ile Ser Ser
                165                 170                 175

Val Ala Ala Asp Glu Phe Gly Gln Ala His Asn Ile Asn Ala Asp Thr
            180                 185                 190

Cys Ala Gly Glu Leu Ala Ala Ala Leu Gly Ala Glu Lys Leu Ile Leu
        195                 200                 205

Leu Thr Asp Thr Arg Gly Ile Leu Arg Asp Tyr Lys Asp Pro Ser Thr
    210                 215                 220

Leu Ile His Lys Leu Asp Ile Gln Gln Ala Arg Glu Leu Ile Gly Ser
225                 230                 235                 240

Gly Ile Val Ala Gly Gly Met Ile Pro Lys Val Thr Cys Cys Val Arg
                245                 250                 255

Ser Leu Ala Gln Gly Val Arg Ala Ala His Ile Leu Asp Gly Arg Leu
            260                 265                 270

Pro His Ala Leu Leu Leu Glu Val Phe Thr Asp Leu Gly Ile Gly Ser
        275                 280                 285
```

Met Ile Val Ala Ser Gly Tyr Asp Leu
290 295

<210> 12
<211> 345
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:CONSENSUS

<400> 12
Met Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
1 5 10 15

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
20 25 30

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
35 40 45

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Ala Xaa Ala Xaa Xaa Xaa
50 55 60

Xaa Xaa Arg Val Asp Val Leu Ser Glu Xaa Leu Pro Phe Ile Gln Xaa
65 70 75 80

Phe Xaa Gly Lys Thr Xaa Val Val Lys Tyr Gly Gly Ala Ala Met Lys
85 90 95

Ser Pro Glu Leu Gln Ala Ser Val Ile Xaa Asp Leu Val Leu Leu Ser
100 105 110

Cys Val Gly Leu Xaa Pro Val Leu Val His Gly Gly Gly Pro Glu Ile
115 120 125

Asn Ser Trp Leu Xaa Arg Xaa Xaa Xaa Xaa Xaa Xaa Phe Arg Xaa Gly
130 135 140

Leu Arg Val Thr Asp Ala Xaa Xaa Met Glu Xaa Val Xaa Met Val Leu
145 150 155 160

Val Xaa Lys Val Asn Lys Xaa Leu Xaa Ser Leu Ile Xaa Xaa Xaa Gly
165 170 175

Xaa Xaa Ala Xaa Xaa Leu Xaa Xaa Xaa Xaa Xaa Xaa Arg Leu Xaa Xaa Xaa
180 185 190

Arg Pro Xaa Pro Xaa Xaa Xaa Xaa Leu Xaa Xaa Val Gly Xaa Val Xaa
195 200 205

Arg Xaa Xaa Xaa Xaa Val Leu Xaa Xaa Xaa Ile Xaa Xaa Xaa His Ile
210 215 220

Pro Val Xaa Xaa Xaa Val Xaa Ala Asp Glu Xaa Gly Gln Xaa Tyr Asn
225 230 235 240

Ile Asn Ala Asp Thr Xaa Ala Gly Glu Xaa Ala Ala Xaa Xaa Gly Ala
245 250 255

```
Glu Lys Leu Xaa Leu Xaa Thr Asp Val Xaa Gly Ile Leu Xaa Asp Arg
           260                 265                 270

Xaa Asp Pro Xaa Ser Leu Val Lys Xaa Xaa Asp Ile Xaa Gly Val Arg
       275                 280                 285

Xaa Met Xaa Xaa Xaa Gly Xaa Val Xaa Gly Gly Met Ile Pro Lys Val
   290                 295                 300

Xaa Cys Cys Val Xaa Xaa Leu Ala Gln Gly Val Xaa Thr Ala Ser Ile
305                 310                 315                 320

Leu Asp Gly Arg Val Pro His Ser Leu Leu Leu Glu Ile Leu Thr Asp
                325                 330                 335

Glu Gly Xaa Gly Thr Met Ile Thr Gly
            340                 345
```